DATA ADMINISTRATION

INPUT

About INPUT

INPUT provides planning information, analysis, and recommendations to managers and executives in the information processing industries. Through market research, technology forecasting, and competitive analysis, INPUT supports client management in making informed decisions. Continuing services are provided to users and vendors of computers, communications, and office products and services.

The company carries out continuous and in-depth research. Working closely with clients on important issues, INPUT's staff members analyze and interpret the research data, then develop recommendations and innovative ideas to meet clients' needs. Clients receive reports, presentations, access to data on which analyses are based, and continuous consulting.

Many of INPUT's professional staff members have nearly 20 years' experience in their areas of specialization. Most have held senior management positions in operations, marketing, or planning. This expertise enables INPUT to supply practical solutions to complex business problems.

Formed in 1974, INPUT has become a leading international planning services firm. Clients include over 100 of the world's largest and most technically advanced companies.

Offices

NORTH AMERICA

Headquarters
1943 Landings Drive
Mountain View, CA
94043
(415) 960-3990
Telex 171407

Detroit
220 East Huron
Suite 209
Ann Arbor, MI 48104
(313) 971-0667

New York
Park 80 Plaza West-1
Saddle Brook, NJ 07662
(201) 368-9471
Telex 134630

Washington, D.C.
11820 Parklawn Drive
Suite 201
Rockville, MD 20852
(301) 231-7350

EUROPE

United Kingdom
INPUT, Ltd.

West Germany
NOVOTRON GmbH
Am Elizabethenbrunnen 1
D-6380 Bad Homburg
West Germany
Telex 418094

ASIA/AUSTRALIA

Japan
Overseas Data Service
Company, Ltd.
Shugetsu Building
No. 12-7 Kita Aoyama
3-Chome Minato-ku
Tokyo, 107
Japan
(03) 400-7090
lex 26487

Ashisuto
-Suzumaru Bldg., 6th Floor
Nishi Shimbashi
me Minato-ku
o, 105, Japan
437-0654
781 26196

pore
ware Consultants (PTE) Ltd.
Pangkor
ore Park
pore 1025
142

INPUT
Planning Services For Management

DATA ADMINISTRATION: EXPERIENCES AND OUTLOOK

JUNE 1984

DATA ADMINISTRATION: EXPERIENCES AND OUTLOOK

CONTENTS

EXHIBITS

1 INTRODUCTION

I INTRODUCTION

A. OBJECTIVE, AUDIENCE, AND NEED

- Data Administration (DA) is becoming more sophisticated. Publicity, confusion, and claims are growing. The objective of this report is to:
  - Clarify what DA is and is not.
  - Delineate the pressures affecting DA.
  - Define related concepts, such as data and information.
  - Provide clear guidelines for:
    - Deciding what specific enterprises would benefit from DA.
    - Technically establishing or strengthening DA.
    - Coping with the human aspects of DA.
- The primary audience is information systems (IS) managers who need both a better understanding of DA and a guidebook to its management.

- The secondary audience is data administrators and data base administrators who need to mesh their operations with broader organizational goals.

- The tertiary audience is both vendors of DA tools and consultants who need a broader understanding of their markets.

B. SCOPE AND METHODOLOGY

- This report is part of the Information Systems Program (ISP) Corporate Systems Planning Program. It addresses the following issues:

  - Current trends and pressures affecting Data Administration (DA) (Chapter III).

  - The theoretical promise of DA versus actual experience with it (Chapter IV).

  - Guidelines for determining what kind of organization can benefit from DA (Chapter V).

  - Functions to be performed in establishing a technically sound DA operation (Chapter VI).

  - Human aspects of DA (Chapter VII).

  - Recommended strategies for using DA (Chapter VIII).

- Information from this report was gathered from the following sources:

  - Interviews with four specialists in DA.

- Interviews with six vendors of DA tools and consulting services.

- Structured interviews with 18 managers in organizations with data base management systems or DA operations.

- Types of organizations included are:

  - Banking, finance, and insurance companies.

  - Commercial manufacturers.

  - Aerospace manufacturers.

  - Large public utility companies.

  - A large state university system.

- The conclusions and recommendations in this report are a synthesis of INPUT's experiences plus:

  - The procedures and results reported by the organizations.

  - The most sensible proposals from the vendors.

  - The most impressive ideas from the researchers.

C. RELATED INPUT REPORTS

- Background and amplification of some of the concepts in this report may be found in five other INPUT reports:

- Micro-to-Mainframe Systems Experiences, May 1984.

  . Will concentrate on the experiences of organizations that use personal-computer-to-mainframe systems. It will also identify systems requirements and project future effects.

- Integrating Systems and Corporate Planning, March 1984.

  . Describes approaches to achieving an integrated information system and corporate business planning process that also achieve full benefits from information technology.

- The Opportunities of Fourth-Generation Languages, September 1983.

  . Analyzes the extent to which fourth-generation languages are used and how they fit into the information systems strategy.

- Organizing the Information Center, August 1983.

  . Discusses how to organize an information center, including chargeback methods.

- Personal Computers in the IS Strategy, December 1982.

  . This report recommends the most effective ways for IS to become involved with personal computers (PCs).

II EXECUTIVE SUMMARY

II EXECUTIVE SUMMARY

- Note: this executive summary is designed in a presentation format in order to:

  - Help the busy reader quickly review key research findings.

  - Provide a ready-to-go executive presentation, complete with a script, to facilitate group communication.

- The key points of the entire report are summarized in Exhibits II-1 through II-6. On the left-hand page facing each exhibit is a script explaining its contents.

A. DATA ADMINISTRATION EXPERIENCE AND OUTLOOK

- This report was produced as part of INPUT's Information Systems Program (ISP).

- Data Administration (DA) is receiving increasing publicity, and for good reason:

  - By improving the accessibility and quality of the data, DA can improve computing on personal computers, etc.

  - The same factors can raise the speed and quality of professional programmers' work.

  - Better understanding of data can also make Management Information Systems' reports more informative. Executive perspectives can be better represented.

- The new report:

  - Distinguishes between DA and related functions such as data base administration.

  - Explains the theoretical promises of DA and cites relevant experiences.

  - Contains guidelines for estimating an enterprise's need for DA.

  - Contains clear "how to" sections on DA.

    . Explains the technical goals that must be met.

    . Gives advice on the human aspects of DA.

  - Offers a strategic guide to the use of this report.

EXHIBIT II-1

DATA ADMINSTRATION EXPERIENCE AND OUTLOOK

- Data Administration's Time Has Come. It Will:
  - Improve End-User Computing
  - Improve Application Programming
  - Move toward Executive Information Systems
- Scope of Report
  - Defines Data Administration
  - Theory and Evidence
  - Guidelines for Estimating Need
  - How to Do It
  - Strategy

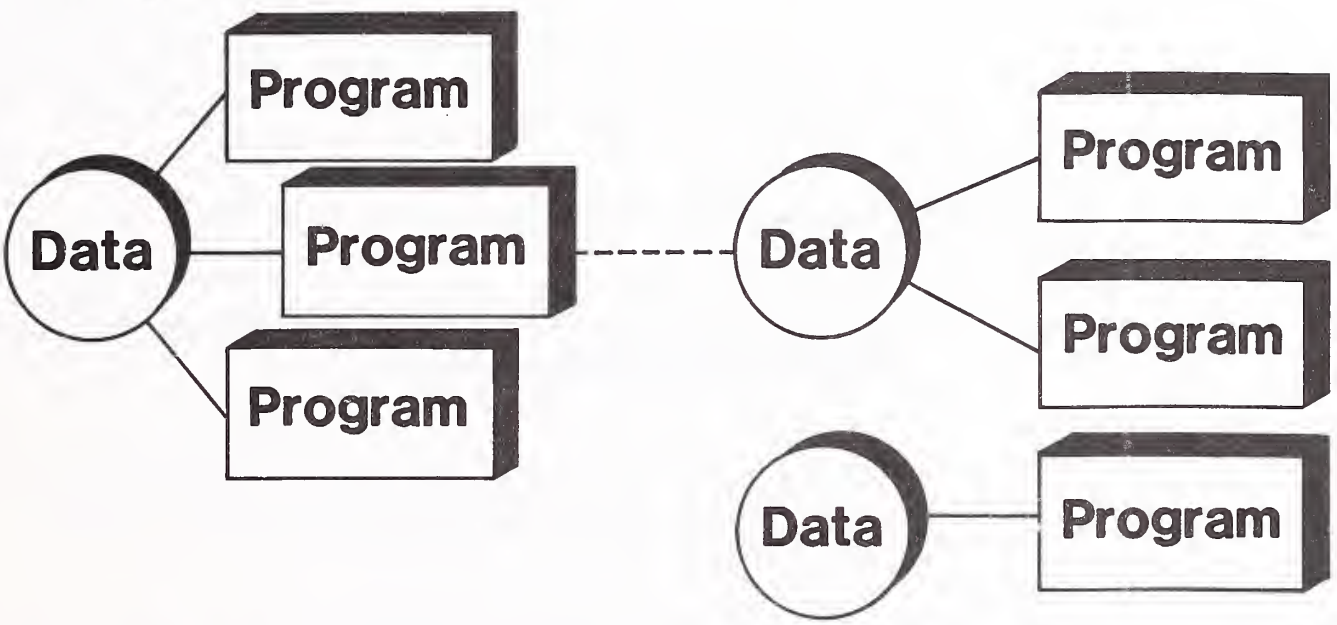
B. FOCUS ON DATA FACILITATES PROGRAMMING

- Conceptually, DA takes a new view of data.
- Traditionally, orientation has been toward computers and programs.
  - A program was written, and often a data base was generated just for that program.
  - For economy's sake, closely related programs tapped an identical data base.
  - A program can feed information to a different data base under certain conditions, but normally a human is required to help this process.
- The orientation of DA is primarily toward the data rather than toward the programs and computers.
- DA seeks to integrate the data bases within an enterprise.
  - Programs can be more readily built on these data.
  - Query languages and fourth-generation languages can more readily be used to "massage" the data.
  - Higher interpretations of the data will offer more meaningful information.
- The trick to all of this is to control the data: to standardize the configuration, map access routes, and control changes.

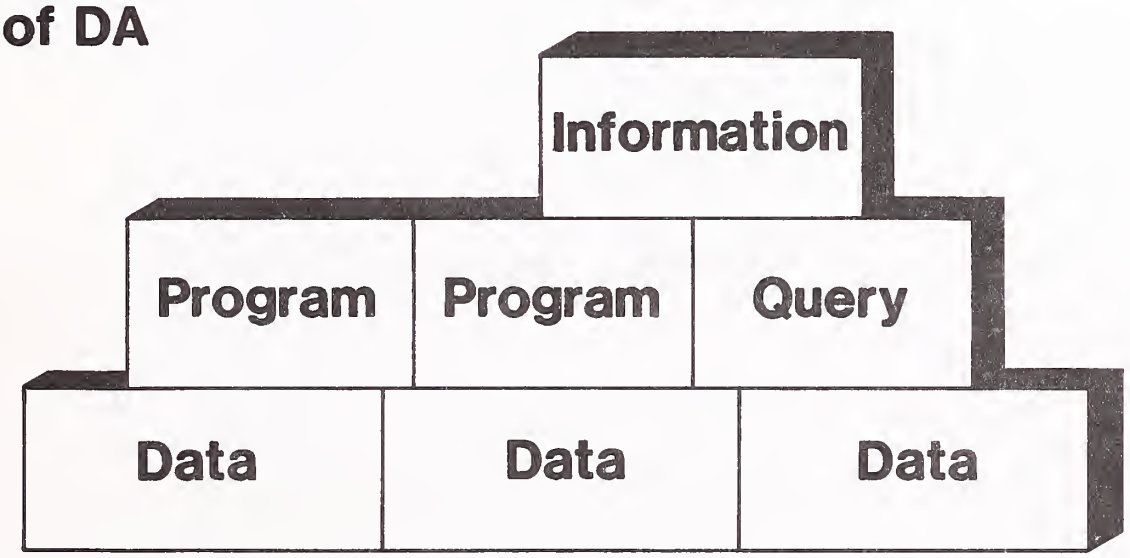
EXHIBIT II-2

FOCUS ON DATA FACILITATES PROGRAMMING

Previous Concept



Concept of DA



C. INFORMATION DEPENDS ON ITS ENVIRONMENT

- "Uncertainty" can be defined mathematically. Information is that which reduces uncertainty.

  - One bit of information reduces uncertainty by one-half.

    . If you are told which half of a football field the ball is on, your uncertainty is reduced by one-half.

    . If a computer printout tells which of two machines is causing more defects, your uncertainty is reduced by one-half.

  - The key to information is the matrix in which it fits.

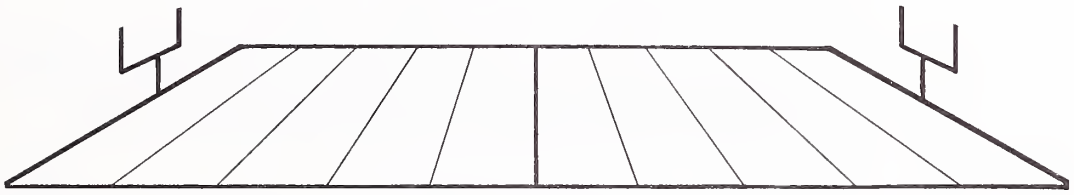
    . Without the concept of the football field, the location of the ball would carry no information.

    . Without the matrix of comparison of the two machines, the data on the computer printout would convey no information.

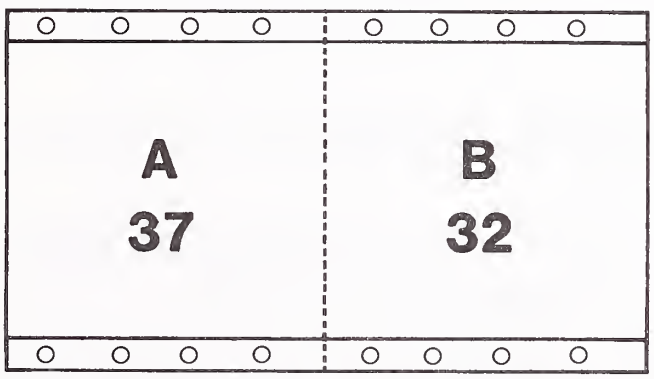
- A major challenge to DA is to see that the data lose no information as they are transferred from one matrix to another and to restructure the data to provide information to the new matrix.

EXHIBIT II-3

INFORMATION DEPENDS ON ITS ENVIRONMENT



On a Football Field



On a Computer Printout

D. DATA ADMINISTRATION BUILDS AN INFRASTRUCTURE

- A serious practical problem in justifying DA is that its existence really demands an infrastructure.

  - An integrated data base is an infrastructure. It is analogous to automated telephone switching equipment or to a network of power lines.

  - Any infrastructure requires a capital investment. This investment can rarely be justified by any single application.

    . An automated switching system for any single residential telephone customer would not make economic sense.

    . A large integrated data base for any single application program would not make economic sense.

  - However, many applications are predicted - some logically and some on faith.

  - Current experience indicates:

    . The predictions of faster application programming and more satisfied end users are valid.

    . Some improvement in management information is reported; e.g., one manufacturing company defined a profitable market.

    . Major manufacturing firms have high ambitions; they want to integrate information to improve the "organicity" of their manufacturing.

EXHIBIT II-4

DATA ADMINISTRATION BUILDS AN INFRASTRUCTURE

- Concept
  - Analogies to Dial Phones
  - Requires Investment
  - No Single Justification
  - Many Predicted
- Current Experience
  - Faster Application Programs
  - Satisfied End Users
  - Marketing Strategies
  - Ambitions in Manufacturing

E. THE CONCEPTUAL DATA MODEL IS THE TECHNOLOGICAL KEY TO DATA ADMINISTRATION

- The technical germ of Data Administration is the "three-scheme architecture."

- Data needed by the enterprise can be organized into a conceptual map called the Conceptual Data Model (CDM).

  - The CDM is described independently of any device that will hold the actual data.

  - The CDM is also described independently of any particular user's view of the data.

  - The CDM is exposed to a vital technical process called "normalization."

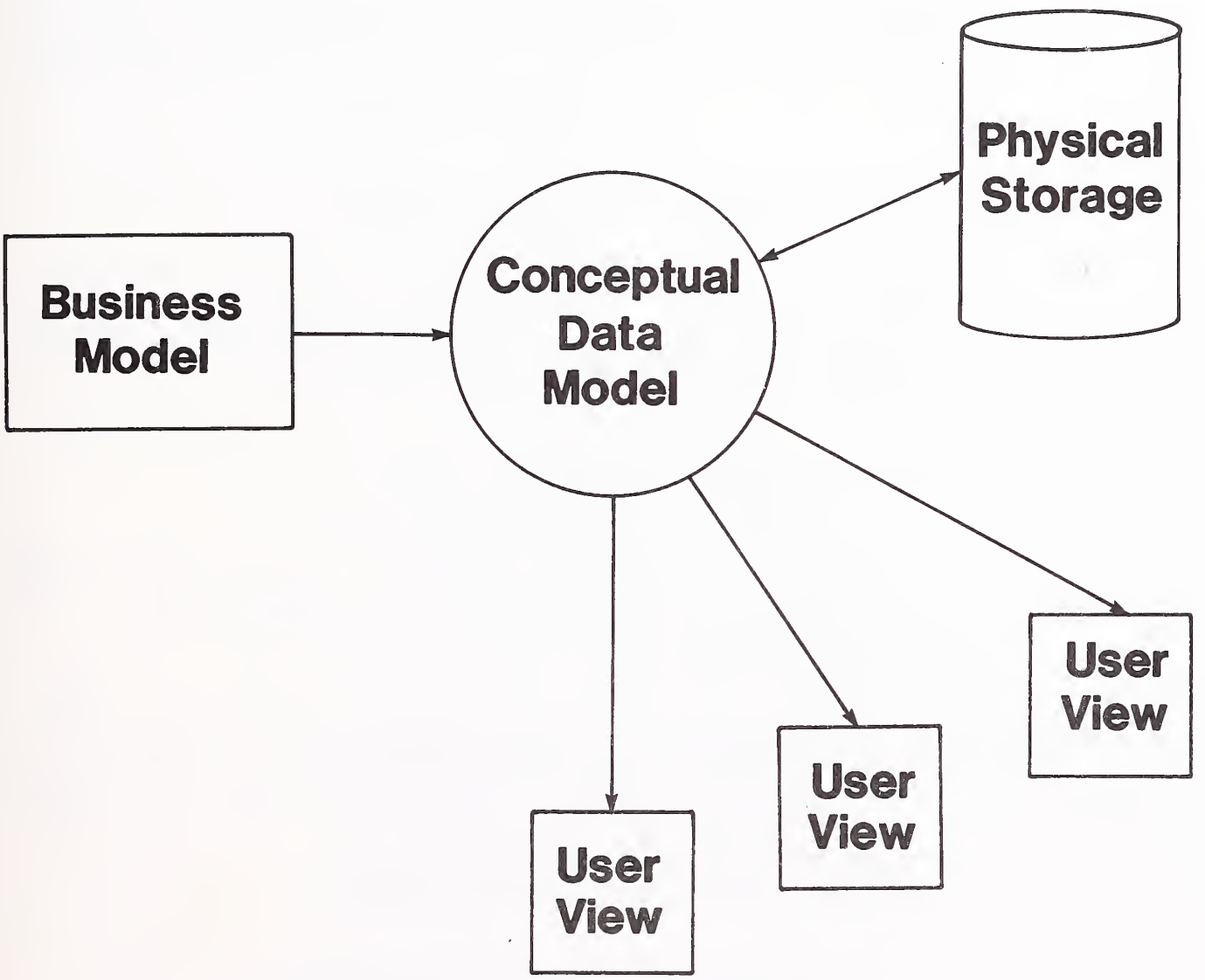
    . Normalization is to the CDM what the alphabet is to a dictionary.

    . Without normalization, a CDM would lack a logical access path.

    . Without normalization, a CDM would lack a framework within which new entries could be classified.

  - A business model is developed to make sure the CDM does contain a framework within which all of the information needed by the enterprise can be represented

THE CONCEPTUAL DATA MODEL IS THE TECHNOLOGICAL KEY TO DATA ADMINISTRATION



F. BARRIERS EXIST, BUT DATA ADMINISTRATION IS RECOMMENDED

- DA is recommended for any organization in which data control is a problem and in which data exploitation is an opportunity.
  - Lack of DA causes problems with data that are redundant, inaccurate, or unavailable.
  - Evidence indicates that DA produces more effective programming and contributes to more informative management information systems (MIS).
- Technical and human barriers stand in the way of DA.
  - The technical barriers can be overcome by generating a conceptual data model, enforcing data standards and utilizing software tools and consultants, where appropriate.
  - Human barriers can be minimized.
    - Rational, understandable explanations can be given to top management whose support is essential to DA's success.
    - Any new technology changes the roles people play in an organization, and the integrated data base technology behind DA is no exception. By understanding the pressures on them, most people can adapt to these changes.

EXHIBIT II-6

BARRIERS EXIST, BUT DATA ADMINISTRATION IS RECOMMENDED

- Lack of DA When Data Are:
  - Redundant
  - Inaccurate
  - Unavailable
- Good DA Results in:
  - Better Programming
  - More Informative MIS
- Overcome Technical Barriers
  - Generate a Conceptual Model
  - Enforce Data Standards
  - Use External Resources
- Minimize Human Barriers
  - Management Understanding Needed
  - Role Changes Required

III TRENDS AND PRESSURES IN DATA ADMINISTRATION

III TRENDS AND PRESSURES IN DATA ADMINISTRATION

A. WHAT IS GOOD DATA ADMINISTRATION?

1. HISTORY AND DEFINITIONS

- Data Administration (DA) is a relatively new and inconsistently defined concept.
  - Data Base Management Systems (DBMSs) appeared in the late 1960s.
  - Data Dictionaries (DDs) emerged in the 1970s.
  - The term DA was not generally used until the 1980s.
    - Now there is an explosion of articles about DA and related concepts.
    - A new journal, Database, concentrates exclusively on data-base-related reports. Widely circulated magazines such as Datamation and Infosystems are trying to define DA and tell their readers where DA should be located in their organizations.
    - Magazine writers discuss DA very positively. They cite its value in "taming . . . uncontrolled data bases" and thereby reducing

both the danger of lost and inaccurate data and the cost of redundant storage.

- There are different different definitions of DA.

  - In the weakest and narrowest definition, DA is merely the "care and feeding" of a DBMS: that is, maintaining its software, supporting applications programmers who use the DBMS, protecting data base integrity and standards, etc.

    . This definition is wrong, according to most writers and almost all the people surveyed by INPUT.

    . This is really a definition of Data Base Administration (DBA).

  - In the most frequent definition, DA is DBA plus broader business functions:

    . DA also keeps higher executives informed of the challenges and opportunities in data management.

    . DA also keeps IS informed of corporate goals and reflects the goals in the data structures.

  - DA also has different tools and powers than does DBA.

    . The DA function would be impotent without a DD. The DD contains metadata (i.e., data about data). The metadata describe the company's basic data, the users, and the data processes and processors.

    . DA holds "the keys to the kingdom" in its control over the DD: that is, DA has the ability to enter the DD, report its contents, and modify them.

- The terms Information Resource Management (IRM) and Data Management are frequently used to cover the same territory as the above definition of DA.

- The strongest definition includes the above plus a proactive responsibility: to derive information that will influence corporate goals, and even help create new ones. INPUT believes this is the proper definition for DA.

- This type of DA needs a business model (described in Chapter VI) to add to its functions.

  . A business model derives from what the business does and not from what happens to be in the business' present data bases.

  . Business models aid in the anticipation of corporate needs for information.

- Exhibit III-1 illustrates these definitions.

• Few organizations have a DA function that meets the strongest definition.

- Only about a quarter of the mainframes now in use support DBMSs. (A higher percentage do have some kind of data base technology.)

- In that quarter, most have a rather narrowly defined DA function.

- Most DBMSs are not used optimally. This opinion is widely held by the INPUT respondents, and especially by the academic researchers and commercial consultants.

EXHIBIT III-1

DEFINITIONS OF DATA ADMINISTRATION

	WEAKEST DEFINITION	COMMON DEFINITION	STRONGEST DEFINITION
Major Results:	Facilitate Application Programming and Data Access and Quality (By Improving Data Bases.)	Facilitate All Programming and Data Access and Quality (By Fostering Common Data Bases for Different Applications.)	Facilitate All Programming, as at Left. Also Improve Corporate Decision Making (By Providing Better Information.)
Major Tasks:	Maintain DBMS, Keep Data Definitions Current and Constant, etc.	Define and Manage Logical Data Structure and Data Transform Standards. Lobby to Expand Common Data Base Use. Begin to Reflect Corporate Goals in Data Base	Continue Tasks at Left. Explain Data-Oriented Development. Identify Enterprises to Be Served and Voids in Needed Information.
Major Tools:	DBMS	DBMS DD	DBMS DD Business Model Modeling Tools

- Representative comments include: ". . . only a few (firms) have truly implemented DA."

- "Advertisements constantly appear for programmers with experience in particular data base products. (But) it is rare to find a succinct description of a data base job dealing primarily with the company's information requirements."

- "Most DBMS users continue to use the same analysis and design methods as they used without DBMS and thus have produced one data base per application instead of multipurpose, shared data bases."

- How many DA functions are operating now? The answer depends on one's definition of DA. Exhibit III-2 shows (1) the estimated number of IS organizations with hardware that could support DA, but which have essentially no DA function; and (2) estimated numbers as a function of how weakly or strongly DA is defined.

- All consultants contacted by INPUT agreed that the DA concept is spreading. As Exhibit III-2 shows, there is much room for growth.

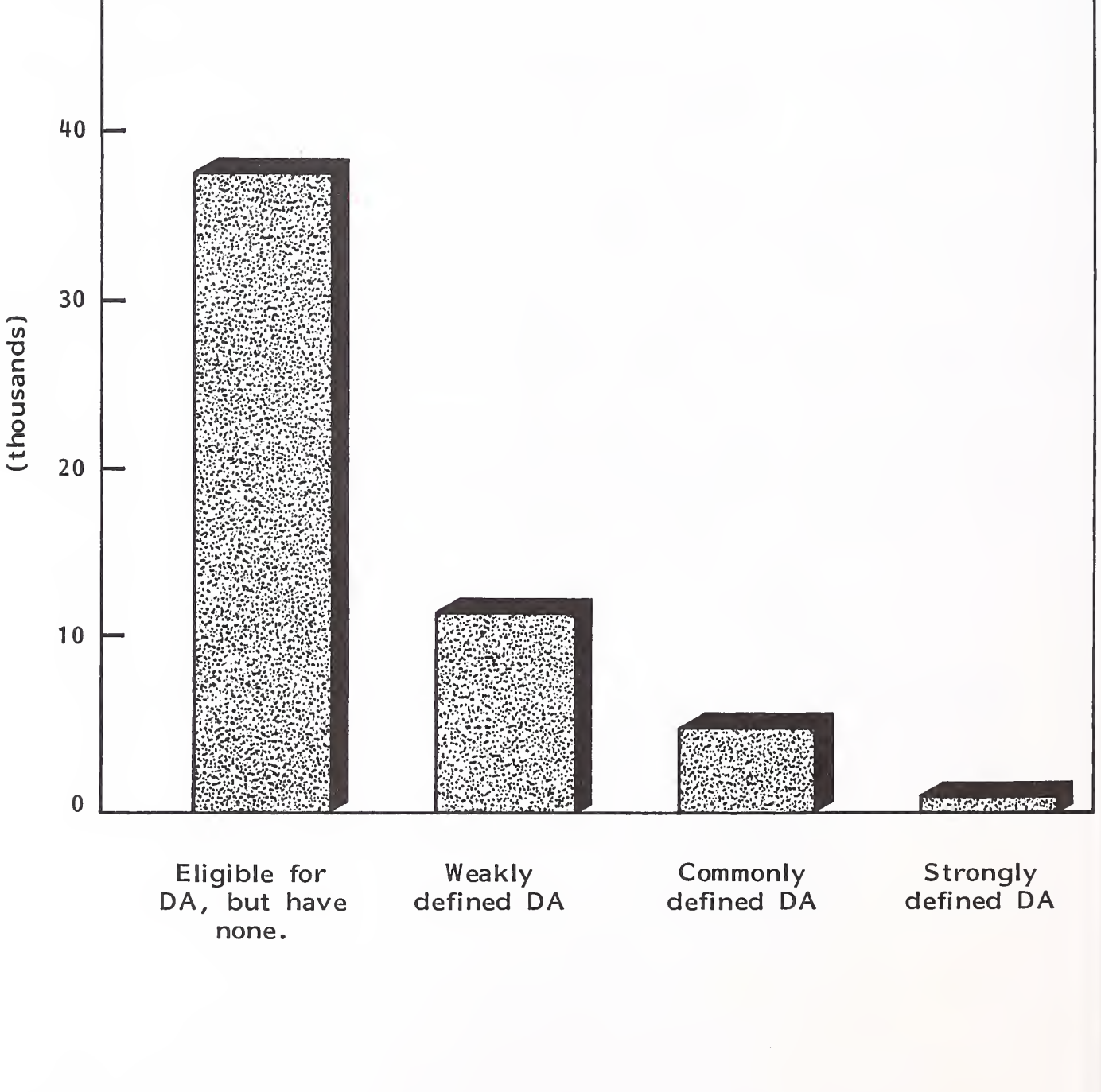
  - Before reaching most of the candidate sites, commonly defined DA could grow by a factor of ten.

  - Strongly defined DA could grow by a factor of 50 before reaching most present candidates.

  - As the population of computers grows, so will the number of locations needing DA.

EXHIBIT III-2

NUMBERS OF ORGANIZATIONS WITH DATA ADMINISTRATION
AS A FUNCTION OF DEFINITION OF DATA ADMINISTRATION



2. MEASURES OF GOOD DATA ADMINISTRATION

- There are agreements, in principal, of what constitutes "good" DA. In general "good" DA can be defined economically.

  - A company's data represent an investment. It is as real as the company's investments in equipment or training.

  - DA is a rational attempt to maximize the return on the investment in data.

- "Good" DA can also be defined pragmatically.

  - DA is good to the extent that it delivers what its users expect and is a help in making them more productive.

  - DA is even better if it also serves as the basis for realistic, top-level overview reports to executives.

B. CURRENT TRENDS AND PROBLEMS

1. TECHNICAL TRENDS

- Users of DA and Information Systems (IS) services are changing. In the past, IS had clearly defined users. Therefore it seemed to have clearly defined data.

  - Now users are more transient. More "amateurs" are in the act because:

    . They want to use their personal computers for personal information delivery.

        - . Decision support systems are proliferating, and users want data for them.

    - Therefore data extraction is becoming more demand driven and less planned.

- The physical environment of DA is changing.

    - Telecommuncations are becoming more readily available.

        - . Local area networks (LANs) increase the volume of data communications.

        - . But LANs do nothing to improve the quality of the data.

    - There is a strong trend toward distributed processing.

    - These physical trends create pressures to serve more remote users.

- The companies' mainframe computers have lost their monopoly on large data bases. External data are increasingly available through commercial, on-line data base services. With the proper micro or terminal:

    - The company marketing department can call up external data bases to review "Sources Sought" in the Commerce Business Daily.

    - The company physician can ask for references to current medical information from a service sanctioned by the Amercian Medical Association.

    - Nonspecialists can interrogate encyclopedic data bases such as DIALOG, BRS, and ORBIT.

- A bank in Massachusetts can now legally exchange masses of data with affiliated banks in Rhode Island, Maine, and Connecticut. There is a legal as well as a technical trend toward regional banking.

2. PUBLICITY AND PROBLEMS

- The above trends generate publicity about the causes of changing user types (e.g., with respect to personal computers and decision support systems), and their effects:

  - The concept of DA (or IRM) is being diffused into the IS culture.

  - Underlying theoretical data structures are seen as stable, whereas applications and technology are constantly changing. So, in a search for stability, some people cling to data rather than technology.

  - Some people see the traditional IS departments as obsolete. The view of these people is affected by publicity about the power of:

    . Decision Support Systems (DSSs)

    . Fourth-Generation Languages (FGLs)

    . Personal Computers (PCs)

- These trends cause problems.

  - True believers in program documentation are appalled: "For thirty years we've failed to get proper documentation out of professional programmers. Now what can we expect out of a bunch of amateurs with PCs?"

- INPUT heard a repeated complaint: "PCs are duplicating data bases all over the company!" Consultant Dan Appleton warns of "information pollution." If only because of PCs, DA has a harder job: It is now more difficult to:

  . Avoid redundancy of data.

  . Ensure security of data.

  . Maintain the currency of data (i.e., make sure no one is using obsolete data.)

- DA can help solve these problems. (Details are in Chapter IV.)

  - Good data structures facilitate the documentation of application programs.

  - Strong DA can control data pollution.

- Another problem arises: Some IS people don't like DA. They perceive DA as a violation of their turf. (See Chapter VII.)

- In summary, current technical trends and publicity:

  - May threaten the traditional IS department.

  - May make IS feel hostile to DA.

  - May create a greater need for DA services.

C. THE NEW PRESSURE: WORK IN, FUN OUT

- A psychological fact will become increasingly obvious.
  - The fun in using a PC or DSS comes mostly from getting the results.
    - PC and DSS advertisements focus primarily on output.
    - When a PC displays interesting results, it is a nice beast that entertains.
  - The work is all on the input side.
    - Data entry is just clerical work.
    - It's frustrating to fuss around manually finding, generating, or reformatting the data.
- This fact (work in, fun out) carries an important corollary: Data will be shared.
  - At the individual level, people will swap data.
    - At this level, "good" data are likely to be defined more by shareability--having a common format--than by accuracy.
    - This unofficial data swapping is already starting to happen. One consultant says: "Because the necessary data bases aren't created, and official numbers aren't established, a black market of data is generated by users. People with their little floppies are now running down the halls saying to each other, 'Here, I've got the good data.'"

. Of the exchanged data, he adds, "Some usually are inaccurate, some of them are outdated, and a lot of them are misleading because of lack of data-naming standards."

- The long history of data bases is summarized in Exhibit III-3.

- Of course the company will still offer data to users. The palatability of this offering will influence the extent of the tendency toward data-swapping by individuals. But the tendency will still exist.

- DA can respond in either of two ways to the grass roots trend toward data sharing.

  - It can fight the trend.

  - It can facilitate it.

- Good DA will recognize the trend and facilitate it.

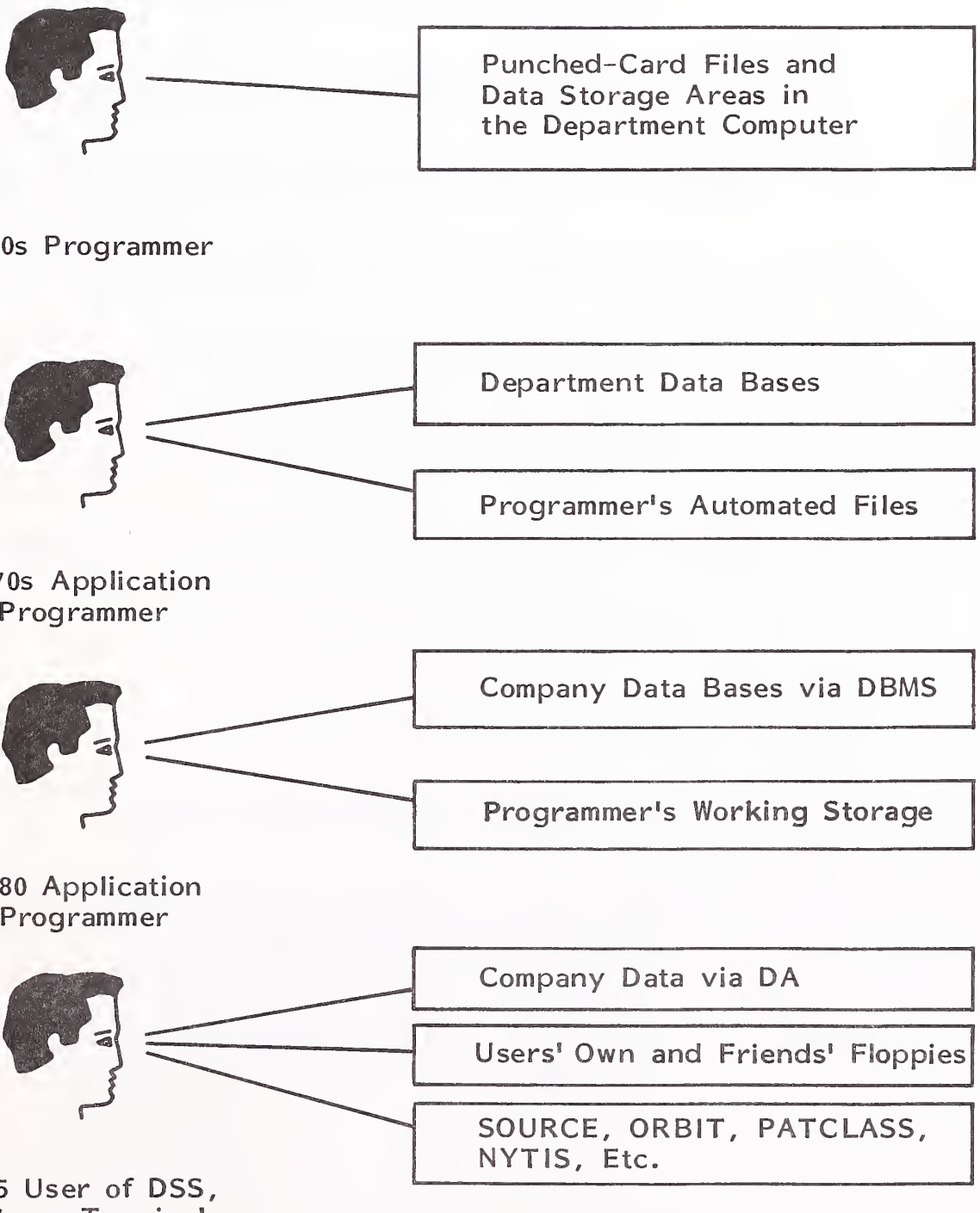
D. LONGER TERM TRENDS

1. TOWARD FOCUS ON DATA

- Harvard Business School's Richard Nolan helped define the DA concept through articles in the Harvard Business Review. In 1979 he traced a typical big business' relationship with a computer through six stages. (Comments about the stages are INPUT's as well as Nolan's.)

EXHIBIT III-3

DATA BASES VISIBLE TO USERS IN DIFFERENT ERAS



- Initiation: The business gets its first computer.

  . Top management pays attention to it.

  . Single-file applications are implemented successfully.

- Contagion: Applications multiply.

  . Top management relegates responsibility to the IS professionals.

  . IS overcommits itself, and its budget booms.

- Control: Focus is first on the budget, then on the technology.

  . Top management fusses about the budget.

  . IS seeks user accountability for expenses.

  . IS considers or acquires a DBMS.

- Data Administration (as identified by the strongest definition).

  . The DA function is formally established.

  . A DD is acquired.

- Integration: Tools and procedures are involved.

  . DBMSs tend to be used first as substitutes for file access methods (e.g., ISAM and VSAM).

  . Newer requests are grouped into functional areas (e.g., finance, manufacturing).

. A need becomes clearer: the identification of common data for these areas.

- Maturity: (not quoting Nolan) The company consciously tries to maximize the return on its data investment.

- A new treatment of Nolan's stages is shown in Exhibit III-4. It emphasizes the common trends:

  - Away from focusing attention on hardware and computer resources.

  - Toward focusing attention on information and data resources.

- This version uses the stages as a scale to measure the IS maturity of individual companies.

- The maturity of IS has presented ever-changing questions to executives. These are outlined in Exhibit III-5.

  - On any scale, the bulk of the industry has yet to reach the "mature" stage.

  - As technology changes, the definition of "mature" will change.

2. TOWARD A DATA UTILITY

- Middle managers have an increasing need for information. (Section III.F defines "information" as that which reduces one's uncertainty about a question; data produce information only when they fit into some kind of "matrix" or frame of reference.)

EXHIBIT III-4

TRANSITION ZONE IN FOCUS OF ATTENTION
ON AN I.S. MATURITY SCALE

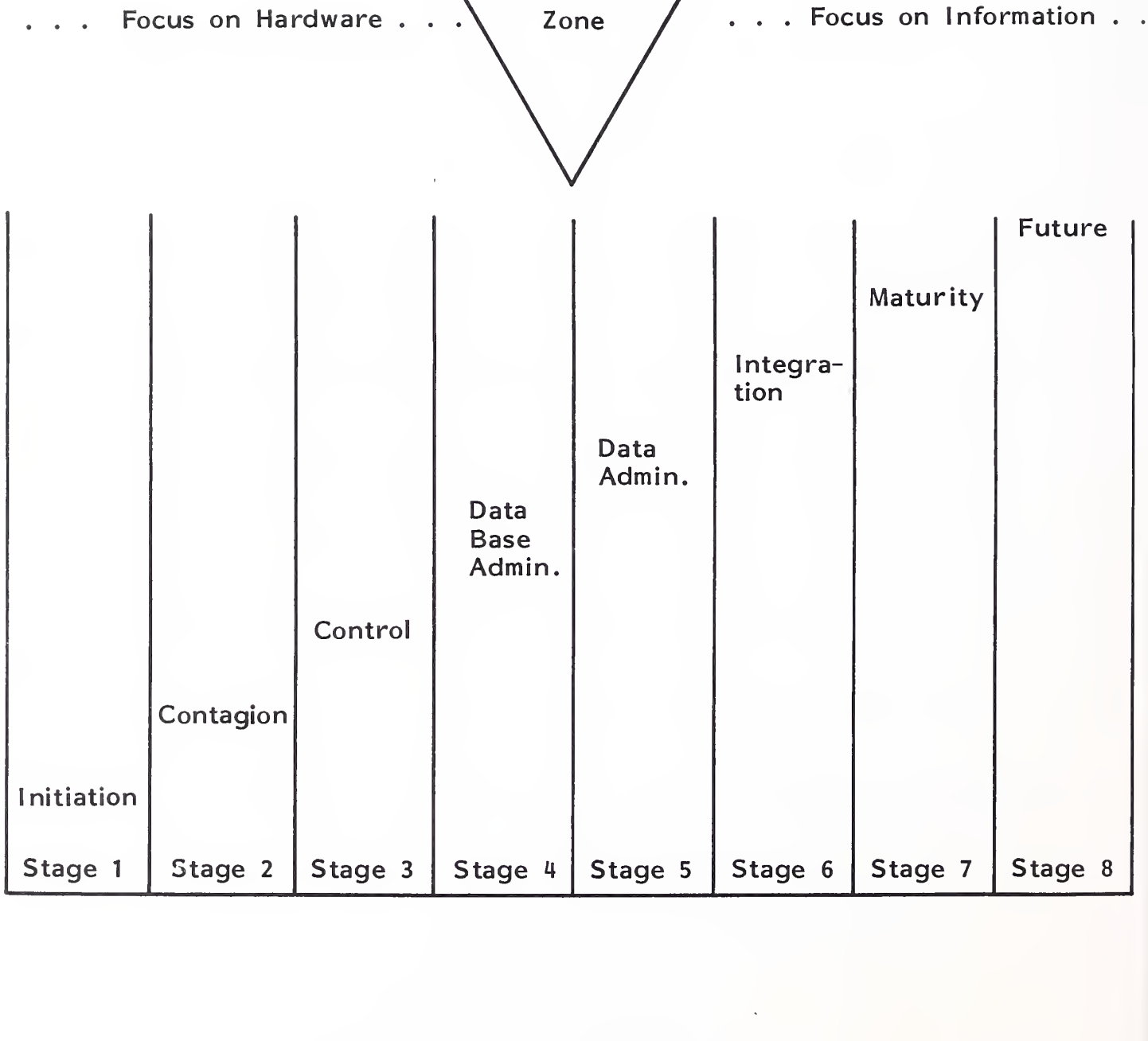


EXHIBIT III-5

QUESTIONS IN THE MINDS OF EXECUTIVES

About 1955:	My accounting and finance people tell me I need to buy a computer. Are they right?
About 1965:	My operations people tell me I need a bigger computer. Are they right?
About 1975:	My IS people tell me I need something called a DBMS. Are they right?
About 1985:	My DA people tell me I need "a data design in which my metadata and entities are normalized to the fifth normal form." Are they right?

- Managers' need for information causes them to want to "massage" data in order to "exercise" different models quickly.

  - American manufacturing is moving toward radically shortened production runs and an increasing variety of products. This trend increases the need for DSSs to answer "What if" questions about new factory configurations and equipment.

  - Financial managers need support in dealing with increased economic volatility.

- "What if" simulations do not directly concern DA. But the data needed for these simulations represent a major DA issue.

- INPUT respondents agreed that "There's a lot of stored data out there" in both corporation and application data bases. There was also agreement on the basic challenge:

  - To control the data (by means of both DDs and policies).

  - To extract data so anyone can use them.

- A trend toward the Information Center concept or toward the establishment of a Data Utility is seen. End users would plug into the Data Utility and create their own reports.

  - A consultant voiced a dissenting opinion: The concept is good, but it doesn't work now. The data is too:

    . Redundant.

    . Hidden.

  - . Out of date or inaccurate.

  - . The data access mechanisms are not user friendly.

- How does one improve the data? By strengthening the DA function as indicated in Chapter VI.

- The popularity of the Information Center concept is a symptom of the outlook for more and more "consumer computing." This outlook creates a dilemma for DA.

  - The pressures for consumer computing cannot be avoided.

  - At present there are grave dangers--consumers cannot generally be prevented from abusing or misusing data.

- What will really happen? Some INPUT respondents see a trend over the next three years toward something like IMS or multiple formats, which feed a repository of DDSs, and which produce something of the character of Lotus spread sheets.

- The industry has little experience with these kinds of information delivery systems. Technically, such systems could be built. But whether they are built or not is basically an executive decision.

E. THE EXECUTIVE VIEW

1. COMPETITIVE FORCES

- IS departments reside in businesses that in turn live in a world that is becoming "smaller" and more competitive. Top executives respond to competitive forces in ways that affect the entire IS department.
  - "Pervasiveness of information gathering . . . (and the) search for focused information" is a characteristic for which Japan is famous. Author Ezra Vogel says the above is probably the most important single element in Japan's continuing economic success.
  - American executives feel the need to make informed, competitive decisions.
  - IS is a logical way to help statisfy that need.
  - IS cannot satisfy the need without a strong DA function.
- For two centuries executives have known that their prosperity comes from:
  - Natural resources.
  - Capital equipment.
  - Labor.
- Now executives know (or at least sense) that their competitive position depends on another factor: Information.

- "Data ownership" is an issue that INPUT discussed with respondents. All respondents had definite opinions.

  . The concensus was well expressed by a director of a large information planning project. He asked rhetorically, "Who owns the data? Who is the custodian? What is the user-creator relationship?"

  . He added, "This is not well understood by users."

- To top executives, data ownership is not an issue. If PCs and floppies contain data that could contribute to better higher level information, then the company owns the data.

- The burden of beating a path to the data falls on DA.

- In companies attempting to "beat a path," the almost-universal mechanisms were:

  - A Data Dictionary.

  - Policies regarding data dictionaries.

2. LEVELS OF DESCRIPTION

- Some of INPUT's respondents mentioned "executive information systems" in addition to the usual term, management information systems (MIS). These respondents recognized that, compared to lower and middle management, top executives need information that is different. Their information should:

  - Cover a broader scope.

  - Be expressed using terms and concepts that executives work with.

- MIS specialists, often in middle management, may not know the executive terms and concepts.

- Software is rarely available to translate raw data into these terms and concepts.

- Exhibit III-6 is a deliberately truncated organization chart. It illustrates the concepts in which each level of management is interested.

  - The organization chart describes a plant that manufactures hybrid electronic microcircuits and three other lines of products. It also has engineering, marketing, financial, and administrative operations.

  - The first-level supervisor needs answers to questions such as: How many hybrids are scheduled through my autobonder today? How many did the autobonder do yesterday? How many were defective? (These questions will be used repeatedly as examples in this report.)

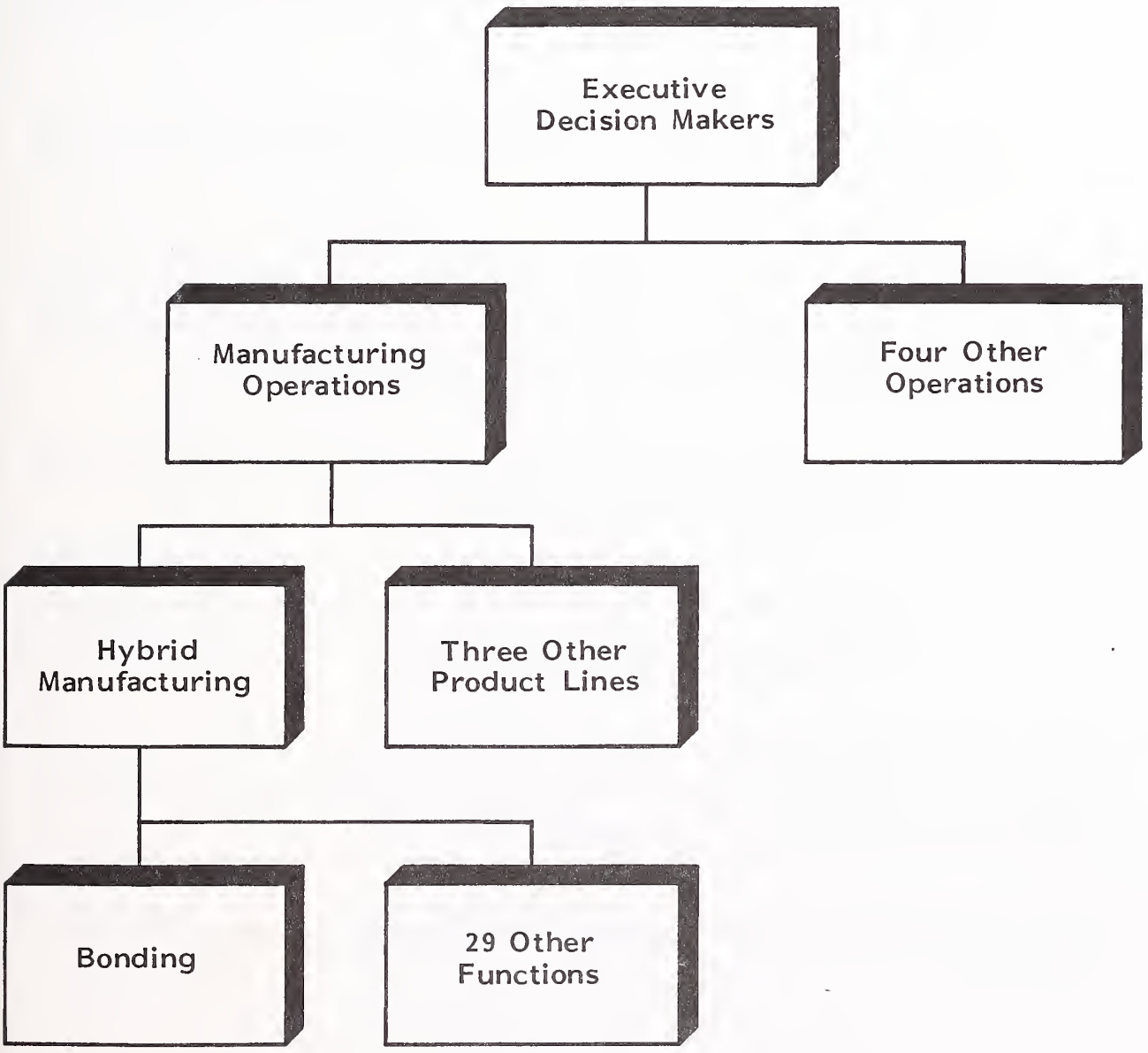
  - The next level of management needs more comparative information, e.g., of the 30 functions in manufacturing a hybrid, where do most defects arise? Which functions are most responsible for our being behind schedule on our contract?

  - Upper-middle management wants to see higher level comparisons: What are the backlogs on our four product lines? Is any product giving us undue problems with defects?

  - Top executives want information to justify high-level decisions. They need to know whether automatic handling and kitting systems for hybrids, or precision milling systems for

EXHIBIT III-6

TRUNCATED ORGANIZATION CHART
TO ILLUSTRATE LEVELS OF DESCRIPTION



inertial products give the greater return on an investment. If the competition could compete in the production of hybrids, how would quality compare?

- The organization chart in Exhibit III-6 is a hierarchy. Answers to questions from different levels on the hierarchy must come from a hierarchy of data.

  - Exhibit III-7 shows such a data hierarchy. It parallels the management hierarchy illustrated in Exhibit III-6.

  - Note that data move up. As they do so, they lose their identity--but not their influence.

    . The first-level supervisor wanted to know how many defects were caused by the autobonder. Such raw data, by themselves, are not needed at the next level of the hierarchy.

    . But the data are necessary for what the next level does need: information on comparative quality.

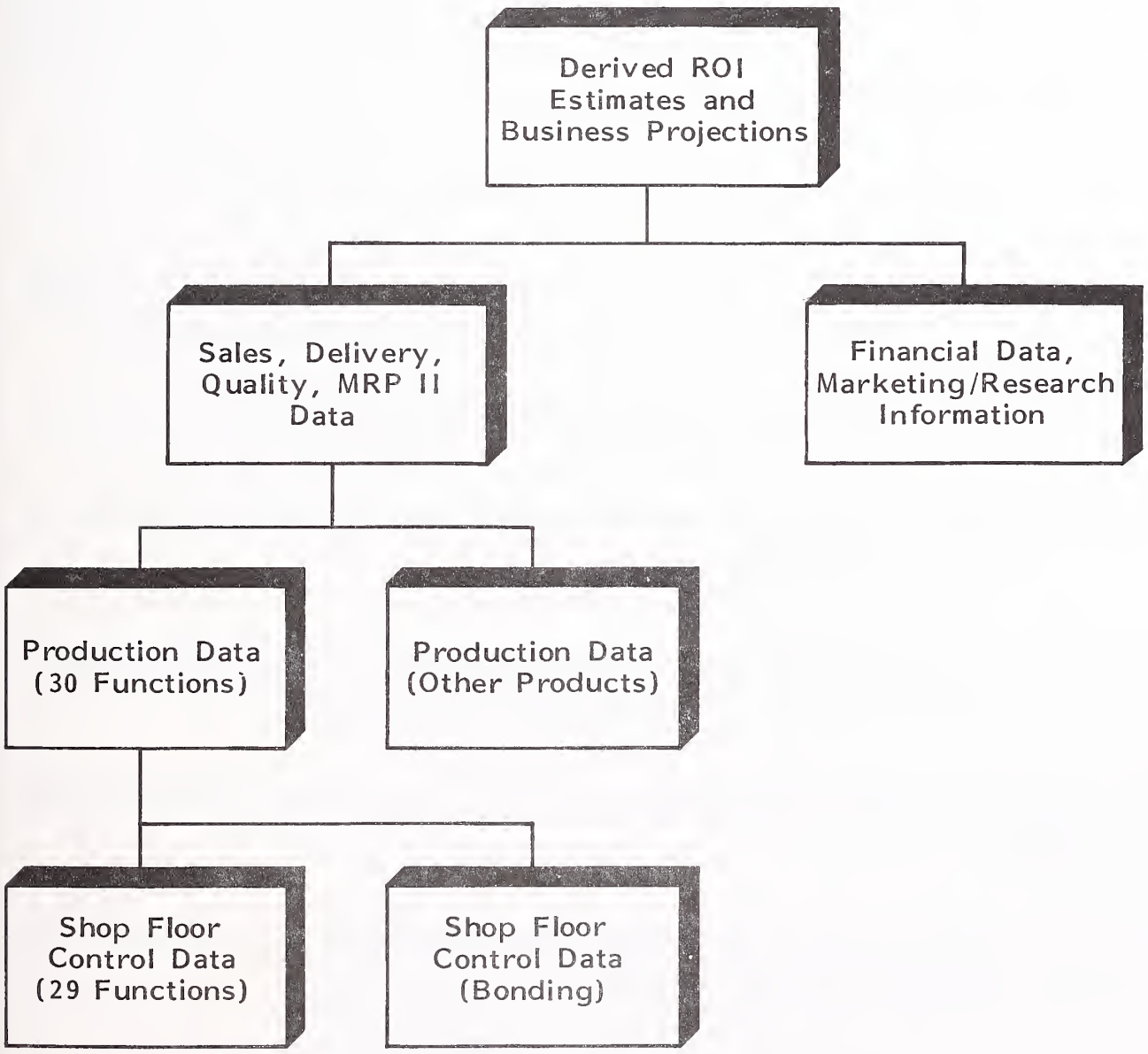
    . Finally, the data exert an indirect influence on the market share estimates at the top of the hierarchy.

- Such a data hierarchy shows why top executives believe all data ultimately belong to the company. All data affecting any company function could ultimately influence top executives' decisions.

- DA's responsibilities, broadly defined, include creating the basis for such a hierarchy.

EXHIBIT III-7

HIERARCHY OF DATA TO
SUPPORT DIFFERENT LEVELS OF MANAGEMENT



F. WHAT "INFORMATION" MEANS

- Many vague definitions of data and information have been presented, with confusing results. The confusion is not necessary. Data and information can be defined precisely.

- Information can be defined mathematically: Information is that which reduces uncertainty. ("Uncertainty" is also defined mathematically, in terms of the homogeneity of a field.)

  - One bit of information reduces uncertainty by one half:

    . Suppose a manufacturing supervisor asks, "Which is damaging more parts--the autobonder or the pick-and-place machine?"

    . The answer conveys one bit of information.

  - Note that the data or answer requires a structure or question before it can convey information.

  - This mathematical definition of information is from Norbert Weiner. It has been used in information theory for forty years.

- Information also can be defined in a half-mathematical, half-psychological manner.

  - One of the first successful artificial intelligence systems, AGILE, was based on adapative matrices.

    . There was once a superstition that the human brain was smooth in the beginning but "added a wrinkle" every time something was learned. This superstition happens to be analogous to what

really happens in an adaptive matrix: Before "learning," the entries in the matrix are "smooth" and uniform. After learning, the matrix entries are more varied.

. Before AGILE "learned," its matrices were homogeneous, and it was therefore "uncertain." As data were converted to information within AGILE, the entries in its matrices created patterns that reduced "uncertainty."

- So, in terms of artificial intelligence, information is that which changes an adaptive matrix.

  - AGILE has been used in connection with adaptive matrices for twenty years.

- Information can be defined psychologically in similar terms: Information is that which changes a mental "matrix."

  - Of course a person's attitudes may prevent the information from "getting through" to make the change.

  - Strictly speaking, information is that which has the potential of changing a mental matrix.

- In short, data convey information after they are processed in an appropriate "matrix."

  - Suppose a person says, "Eight is the factor by which we can speed up coding if we go to NOMAD." Or another says, "Thirty-seven was the number of defects from the autobonder."

    . "Eight" is a datum that conveys little information if it does not go to a mind that knows what NOMAD is, and what the alternative is.

- . "Thirty-seven" gives little information to a "matrix" that does not have a "structure" representing autobonders.

- Exhibit III-8 shows the analogy between:

  - . AGILE, in which data vectors and matrices are literally multiplied as information is relayed.

  - . People, who process data any time that information is relayed.

- Data, then, are numbers or other symbols that transmit information to a compatible "matrix."

- To an information theorist, DA's charter is to create compatible matrices.

- The problem is that the matrices are in different places. In the autobonding example information is carried to the top executive not only by the data, but also by the:

  - Structure of the shop floor control program.

  - Structure of the material requirements planning (MRP) or manufacturing resource planning (MRP II) programs.

  - Structure of the cost planning and evalution system.

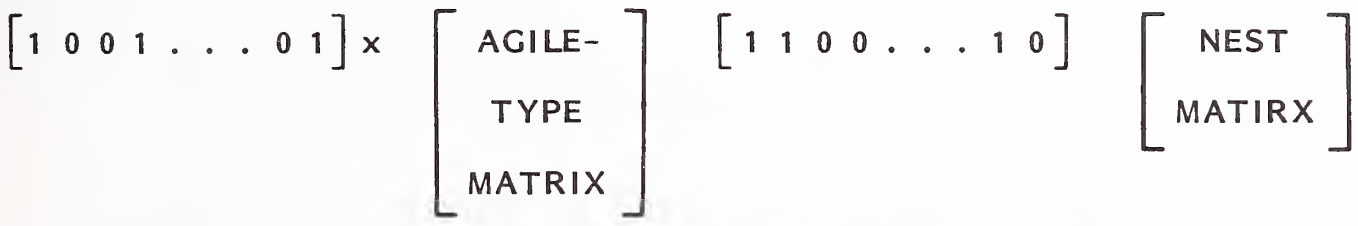
  - Structure of the forecasting and modeling systems.

  - Structure of the data base(s) for all of the above.

  - Documentation on all of the above.

EXHIBIT III-8

DATA AND INFORMATION IN ARTIFICIAL INTELLIGENCE AND IN PEOPLE



Input Data Vector	Matrix in which Data and Structure Combine to Form Information	Output Data Vector	Matrix in which Data and Structure Combine to Form Information

Input Data	"Matrix" in Which Data May Lead to Information	Output Data	"Matrix" in Which Data May Lead to Information

- And, probably, individuals' memories of undocumented peculiarities in the systems.

- Parenthetically, those "memories of undocumented peculiarities" represent one of the biggest challenges to DA.

  - Many IS organizations have some employees whose job security is enhanced by their unique knowledge of the intricacies of some application system. Ideally, that knowledge should be in a "matrix" that is more publicly available and permanent than an employee's mind.

  - DA can strive to minimize these "undocumented peculiarities" by:

    . Strict enforcement of standards, either manually or automatically through a DD.

    . Moving toward data-orientated development, in which more application programs are developed with self-documenting fourth-generation languages.

- As data and information flow to higher levels, more comparisons must be made. For example, information on defective company products must be compared with comparable information about the competitors' products. This information (provided by the marketing department or a consultant such as INPUT) will be in ordinary English. This may not be immediately comparable to the IS reports.

- It is not surprising that most DA managers have found it more rewarding to concentrate on serving (and controlling) their DP community, rather than on trying to serve the top executive directly. One should not apologize for serving the lower levels. They indirectly serve the top.

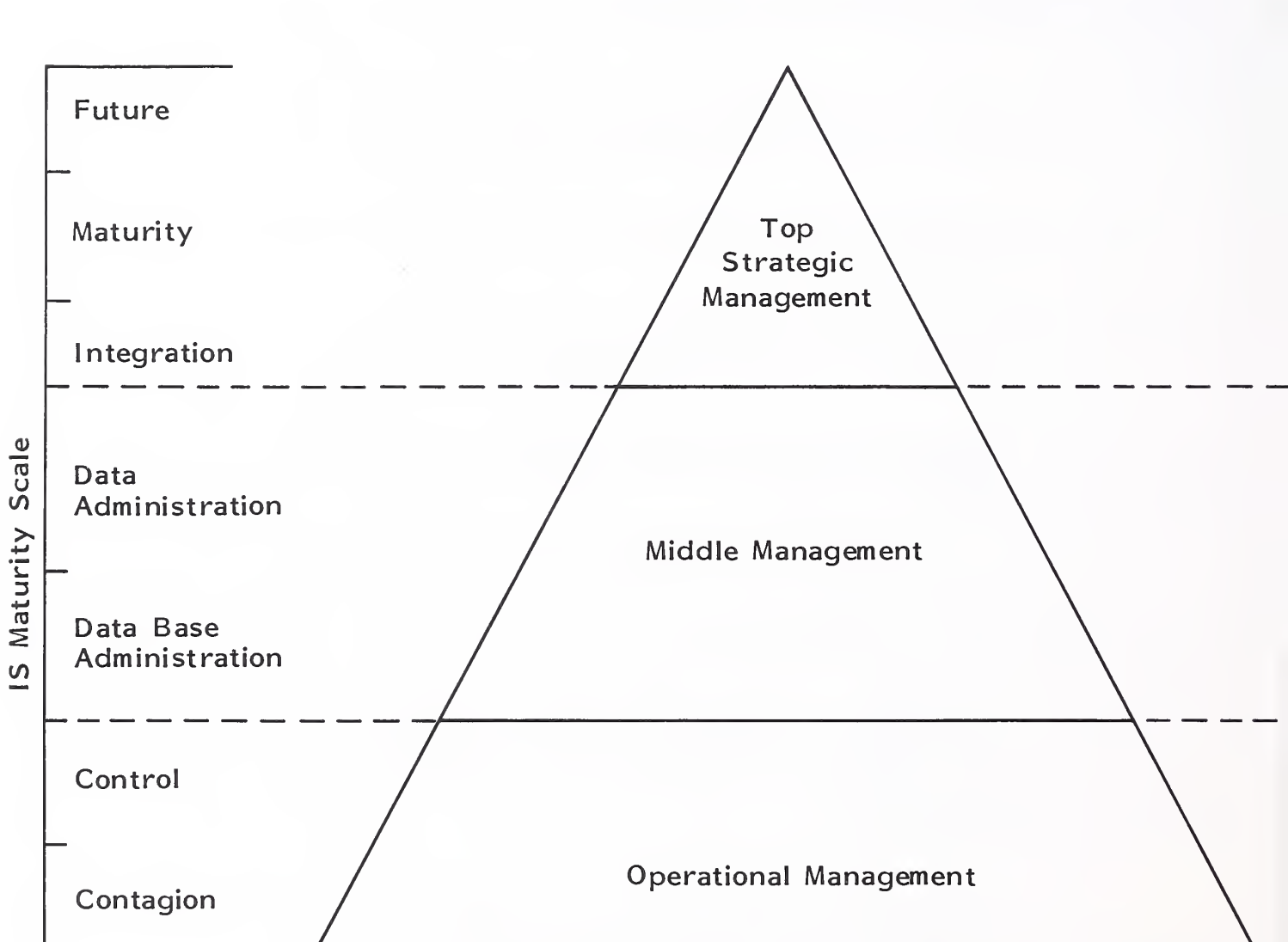
- Nevertheless, the trend is toward IS and DA directly serving higher and higher levels of the organization. This is illustrated in Exhibit III-9, in which a classic organization pyramid is superimposed on the DA maturity scale depicted in Exhibit III-4.

G. PRESSURES FOR CENTRALIZATION

- One of INPUT's respondents defined DA in these terms:
  - "At the corporate level, create and maintain a data model of the enterprise."
  - "Manage the tools to support the data model."
  - "Manage the tools to cause the data structure to conform to the model."
- In this definition, DA is a highly centralized function. There are strong arguments for such centralization.
  - Centralization would satisfy the chief executive by advancing DA along the IS maturity scale to at least the integration stage as indicated in Exhibit III-7.
  - Assuming that the discipline and "the tools to cause the data structure to conform" were strong enough, centralization would:
    - Reduce the black market in questionable data.
    - Let users take advantage of PCs, fourth-generation languages (FGLs), and decision support systems (DSSs).

EXHIBIT III-9

LEVELS OF MANAGEMENT SERVED DIRECTLY
AS DATA ADMINISTRATION MATURITY INCREASES



- Centralization would increase the ability and speed of the corporation in responding to quick questions from customers or government.

- There are also practical pressures against centralization:

  - Creation of the data model can be an enormous job, a costly investment with delayed and uncertain payoffs.

  - Any centralized function may fail to be responsive to the needs of its more remote users.

  - There is a trend toward decentralization of computing. DA goes against this trend.

- Consensus opinons of INPUT's respondents were:

  - "There's value in the concept. But, is it ever really achieved? It's an overwhelming task to set up."

  - "It's a lot of work, a lot of effort. It makes sense on a conceptual level, but it needs to be converted to practice. We need more practical experience."

  - "I'm not sure I see a DA process that is a central management of all the data. Does a central facility clog the flow?"

- INPUT concludes that centralized DA is needed, but only after the enterprise it serves has been defined. This definition is the main subject of Chapter V. The next chapter covers the centralization issue in a review of DA theory and results.

IV DATA ADMINISTRATION THEORY AND EXPERIENCE

IV DATA ADMINISTRATION THEORY AND EXPERIENCE

A. THE INFRASTRUCTURE ARGUMENT

1. BACKGROUND AND SYMPTOMS

- The 1970s saw significant developments in both programming and data base technologies.
  - Large-scale data base technology came of age.
  - So did structured programming.
  - Large-scale magnetic data storage media were already available.
- What did these technologies produce? Generally, they produced "islands of technology" on which each application "did its own thing" without real coordination or integration with the others.
- In such a situation, it is hard to integrate information from one source with information from another source.
  - To service information requests from the management and executive levels, IS typically has to access several application data bases and synthesize the results.

- This synthesis requires expensive and slow "bridge-building" operations between the application islands.

- Meantime, new versions of the request are likely to arrive as management receives the first answers, does not like them, and re-thinks its questions.

- Thus it is difficult for IS to answer high-level questions.

- Lack of integration produces a long list of additional symptoms of an underlying lack of integration:

  - Poor or inconsistent data quality, leading to unreliable information.

  - Inability to combine data into new information structures without first changing data base definitions or programs.

  - Data redundancy.

  - Frequent failure to find data.

  - Inability to share data without manually transcribing or transforming them.

  - Poor data security.

  - Unsatisfactory computer performance.

    . Long search times.

    . Slow response times at individual workstations.

- Poor management of the growth in computer use.

  . Visible project backlogs.

  . Unconstrained introduction of PCs into the workplace.

- Escalating cost of providing the computer power for existing applications, especially in maintenance, where most of the requests for change are serviced.

- Rising discontent with the effectiveness of the computing support provided to the enterprise as a whole.

- Of course, lack of integration is not the only possible cause of such complaints. But it is a factor. And IS management (as surveyed by INPUT) unanimously advocates some kind of overall strategy or mechanism for swift communication between the "islands."

- The highest priority targets of integration strategies are:

  - Data integrability.

  - Data quality.

  - Data accessibility.

2. JUSTIFICATION SKEWING

- One INPUT respondent pointed out a frequent difficulty: one might like to automate a small manual function, but the automation might not be feasible because of the changes that would have to be made in data bases.

- For example, a health maintenance organization is expanding to dental services, and it wants to automate a special credit check on patients receiving major dental treatments. The programming is simple. The problem is that a customer number would have to be changed, and this would entail changes in 35 data bases. This is too much work. Furthermore, there is a danger that not all the necessary data bases will be changed.

- The general situation is this: lacking DA, data bases are not integrated; small enhancements are likely to cause big fan-out problems in maintenance.

- The result is what one respondent called "justification skewing."

  - Large improvements might make sense, but they are infrequent simply because they are expensive.

  - Because the fan-out problems outweigh the advantages, little improvements are not made.

3. INFRASTRUCTURE ANALOGIES

- There is an analogy between phone and electric companies and integrated data bases.

  - If the phone company, in deciding whether to develop automated equipment, had tried to justify automation for each station, we'd still be going through operators for all of our phone calls.

  - If the electric power companies had tried to justify a network of power lines on the basis of a single home or business, we'd all still be using candles and kerosene lamps.

- If large IS departments look for single applications to justify the development of integrated data bases, integrated data bases will not be developed.

- An integrated data base, then, is like a network of power lines or telephones: it is an infrastructure.

  - Needed applications may be built on an integrated data bases.

  - Applications that are discovered to be needed in the future may be built on the integrated data base with today's technology.

  - As new technologies develop, the current infrastructure may support them, too.

B. LOGICAL APPLICATIONS OF THE INFRASTRUCTURE

- Strong logical arguments can be made for infrastructures.

- Most fundamentally, the infrastructure would foster "data-oriented development" versus "application-oriented development," as illustrated in Exhibit IV-1.

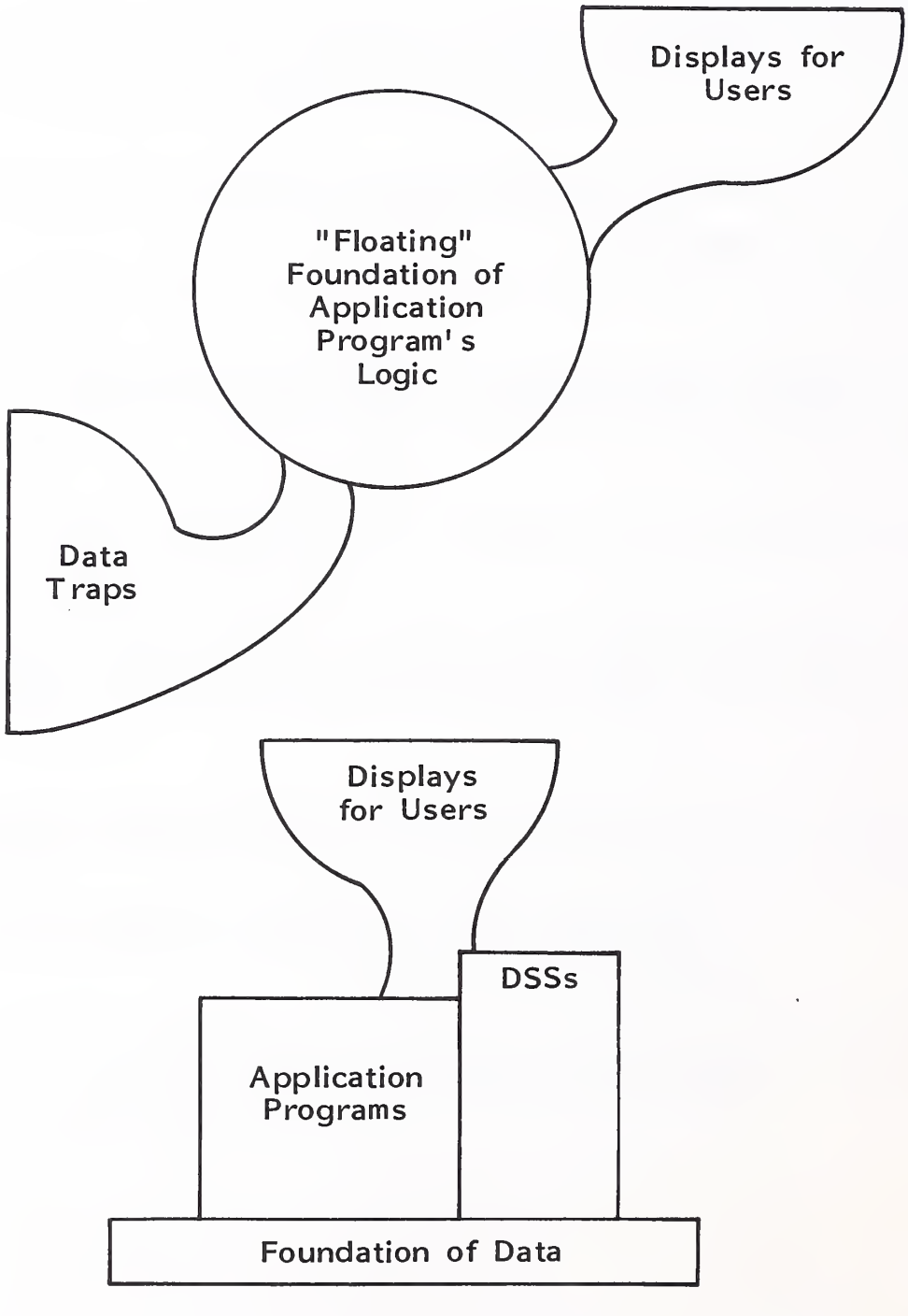
  - The data orientation builds on a solid foundation of data.

  - The application orientation builds on the logic of the program, which is "floating" and must trap data.

  - The cost of trapping the data can be prohibitive.

EXHIBIT IV-1

APPLICATION-ORIENTED DEVELOPMENT (TOP) VERSUS
DATA-ORIENTED DEVELOPMENT (BOTTOM)



I. PAYOFFS TO I.S. AND END USERS

- With infrastructures, end users would find it easier to use FGLs and DSSs because the data to feed them would be easier to find and use.

- IS would find it easier to maintain existing application programs.

  - Many application enhancements affect data bases. With more-integrated data bases, these enhancements would be easier to make.

  - As data bases are changed (e.g., in going from a ZIP code to ZIP + 4), application programs have to be changed. With the support of a strong DA function, the data bases would be more expandable and more capable of transmitting such changes to the application programs.

- Problems with redundant data would be lessened.

  - Storage costs would be cut.

  - Maintenance work would be lower.

  - Security would be easier to accomplish.

  - Most important, currency and version problems would be minimized.

- IS would find it easier to develop new application systems, because they would be based on more standard, available data.

- Auditability would be improved.

- Manual movement of data could be more easily automated.

- Even in a manufacturing operation, it is estimated that 12% of gross sales is spent in moving information. Most of this movement is manual. If half of the movement could be automated, this would double the profit margins of many companies. Given the "infrastructure" of an integrated data base system, the cost of much of this automation would be negligible.

2. PAYOFFS TO MANAGEMENT AND EXECUTIVES

- Data would carry more information to all levels of management. Less information would be lost in nonstandard and undocumented "matrices" as the data move up the hierarchy.

- Data could be translated into tactical and strategic information.

  - Managers would receive more understandable information.

  - Top executives would have higher-level views of the decisions facing them.

- Therefore the quality of decision making should rise.

C. ACTUAL EXPERIENCE

- The DA concept is only a few years old. But implementation of the concept is slow.

  - Carol Shulman, director of marketing for Data base Design, Inc., says, "It takes several years of orchestrated development to achieve shared, nonredundant, stable data bases."

- While a small manufacturing facility built a single shared data base in about two years, a large aerospace corporation plans to spend five years in creating a set of large, shared data bases.

- As a result of the youth of the DA concept and the length of time it takes to reach maturity, complete case histories are almost nonexistent. Quoting one of INPUT's respondents, "We're just now on the brink of some success stories."

- Even where limited case histories are available, dollar savings are hard to quantify.

  - DA offers avoidance of future costs, but it is hard to say what future costs would have been without DA.

    . Business conditions and labor rates change.

    . Technology changes: New software tools and different hardware become available.

  - In short, savings have to be estimated in terms of cost avoidance against a moving target.

- Nevertheless, some clear indications of success are reported. In a finance company, end users can now tap an integrated data base and use FGLs to rapidly generate understandable budget reports.

- A large public utility (with annual sales of more than half a billion dollars) is about a third of the way through its schedule to implement a sophisticated information resource management system. It reports qualitative results and a quantitative prediction.

  - Qualitatively, off-the-shelf application packages are now easier to install, because they can be more easily integrated with other business sytems.

- "Much" data redundancy (and the accompanying threats to data control and integrity) are eliminated.

- Good data structures facilitate a higher quality of application development.

- In terms of speed, the quantitative prediction is this: Within three years, the utility will be developing new applications that are "orders of magnitude" faster than would have been possible in the past.

- Another utility operates a nuclear power plant. The Nuclear Regulatory Commission urgently requested engineering drawings from the utility so that a potential safety problem could be analyzed. The utility had an integrated data base designed in part to facilitate responding to such ad hoc requests. It was able to retrieve the information for the drawings rapidly. Thus it avoided a potential shutdown, which could have cost millions of dollars.

- A manufacturing company developed a facility for searching data bases that had been separate. It was able to identify large products that were sold in the past and that would soon need replacement parts. Thus it was able to satisfy its customers at a high profit margin.

- One of the nation's largest insurance companies has been building a strong DA function for six years. DA is highly centralized, while DP is widely distributed through more than a dozen subsidiary companies.

  . The DA manager believes substantial development costs have been avoided, but he cannot estimate them precisely because of the lack of a stable base of comparison, and because of technical changes (e.g., new languages, etc.)

- . An information center has been created and is working well.

- There are also reports of problems in companies without a DA function or an integrated data base approach.

  - The vice president of a manufacturing company asked for vital cost figures on a product. Three contradictory answers came back from three different data bases.

  - Another manufacturer found that a programmer had set up a system so that a customer had to exist on a file before an order could be taken. So, in the absence of a business model identifying the information that should and should not be needed, a programmer was inadvertently determining company policy on sales!

- Theoretical expectations are compared with the above actual experiences in Exhibit IV-2.

D. PROJECTS IN PROGRESS

- American manufacturing is showing strong, broad interest in integrated data bases.

  - The Air Force is sponsoring Boeing in the development of an Integrated Sheet Metal Center (ISMC).

    - . A major feature of the ISMC is a set of software tools (to aid in design and management) that will be integrated through a common data model.

EXHIBIT IV-2

EXPERIENCE FROM STRONG DATA ADMINISTRATION FUNCTIONS COMPARED TO EXPECTATIONS

THEORETICAL EXPECTATIONS	ACTUAL EXPERIENCE
• Better Use of FGLs and DSSs	• Users Report Fast, Understandable Results
• Easier and Better Application Development	• To Be "Orders of Magnitude" Faster • "Higher Quality" Applications • Purchased Packages Easier to Install
• Fewer Problems with Redundant Data	• "Much" Redundancy Eliminated
• Better Management and Executive Information Systems	• New Markets Identified • Serious Safety Problems Quickly Resolved

  - Another feature is the integration of data from different data bases to coordinate operations all across the factory.

- A major electronics company is in the process of acquiring an integrated system to control the manufacture of microcircuits.

- A major automobile manufacturer is dealing with this problem: How do you keep engineering data in electronic format as long as possible? How do you get information from an engineering data base (perhaps on an IBM computer) to a manufacturing data base (perhaps on a Hewlett-Packard)?

  - It is undesirable to transfer the data to paper along the way because of the cost.

  - Also, the transfer to paper creates opportunities for human error.

  - Finally, the computer-paper-computer route raises the danger that information may be lost as the "matrix" changes.

- The Air Force is sponsoring a study of ways to solve the same problem. McDonnell-Douglas is the prime contractor. Results are expected to be applied in both the aerospace and commercial manufacturing industries.

V GUIDELINES: WHO REALLY NEEDS DATA ADMINISTRATION

V GUIDELINES: WHO REALLY NEEDS DATA ADMINISTRATION?

- This chapter addresses some vital questions.
  - Establishing DA is an effort. What is the scope of that effort?
  - What is the enterprise that DA will serve?
  - What kinds of enterprises have the greatest and smallest real need for DA?

A. DECIDING WHO YOU ARE

- DA cannot be established or augmented intelligently until one identifies what or whom DA serves.
  - As an infrastructure, DA does not merely serve individuals. It serves communities. These communities include:
    - The IS community.
    - People doing "consumer computing."
    - Higher managers.

- These communities form a larger community that DA serves. That larger community is the "enterprise."

- One question is the key to identifying the members' enterprise that DA serves: To what extent are their fates entwined? If consumer computers and the IS community are both supporting the same higher management and no one else and if that higher management has few responsibilities that do not concern the IS community and end users, then all of their fates are entwined.

- If people's fates are entwined, then they are economically integrated.

- The hallmark of DA is integration, and it should serve an integrated enterprise. Before trying to establish or strengthen DA, one should define that enterprise.

- The enterprise certainly need not be a profit and loss center. Typically it is just a well-defined operation within a corporation.

- As a practical matter, one should normally define the enterprise to balance three criteria:

  - The enterprise should be as integrated as possible. It should make economic sense.

  - It may have to include DA's existing responsibilities: DA probably will not have carte blanche to redraw the corporation's organization chart.

  - The enterprise should be small enough to be tractable but large enough to justify the investment in the infrastructure.

  - As a rough rule of thumb, an enterprise using an HP-3000 or larger computer is large enough to justify the investment.

- If it is "above critical mass" economically, a small enterprise is preferable because:
  - Investments in frontend analysis are minimized.
  - DA has a chance to learn.
  - DA has a chance to prove its value before expanding.
- In summary, DA should seek first to serve a small interdependent enterprise.

B. CRITERIA FOR NEED

- The degree to which an enterprise needs DA is a function of several criteria:
  - The volatility of the enterprise.
  - Its size.
  - Its interactivity.
  - Its data orientation.

1. VOLATILITY

- After deregulation, banks and airlines need DA more than they did before. Their data bases are changing more rapidly.
  - Fare structures and routes, for example, are changed more quickly because deregulation allows competitors to change those variables on almost a whimsical basis.

- Interest rates charged and paid, and types of loans are subject to greater change because the competition is freer.

- Such changes increase the need for flexible, expandable data bases.

- In general, the more volatile the enterprise, the more it needs DA.

2. SIZE

- When more interactions are needed among data bases, DA is needed more.

  - The number of interactions varies with size. Within some limits, any doubling of the size of the enterprise will roughly quadruple the number of necessary interactions.

  - Also, as size increases, human memories and personal acquaintances become poorer and poorer substitutes for a DD.

  - In the organizations surveyed by INPUT, the larger ones tended to have more mature DA functions.

- In general, then, the larger the enterprise, the more it needs DA.

- But at some point this need is balanced by the practical advantages of dealing with a small enterprise.

3. INTERACTIVITY

- Some organizations inherently have more in-house interactions than others.

  - An example of an interactive company is a firm that sells computer systems to large dental offices.

- Training the customers is a problem for the company.

- As application programs are changed, the trainers need to get information about the changes.

- As the trainers have problems, the sales department needs to know about them.

- The manufacturing department buys other companies' computers and peripherals, and lashes them together into a system with the firm's own software. As sales change, the firm will buy from different hardware suppliers.

- The application programming department needs to know about hardware changes.

- As application programs change, the trainers are affected.

- A company in the same size range is an accounting service firm. But it is not very interactive. It simply provides clients with accountants on an overload basis.

- The more intrinsically interactive companies have more interactions among their data bases and therefore have a greater need for DA.

4. DATA ORIENTATION

- Some companies are more oriented to masses of data than others. Even if a parts distribution company and a mining company are exactly the same size in terms of gross sales.

- The distributorship may have ten times as many employees as the mining company.

- The distributing company may have a thousand times as many customers.

- The distributorship therefore would have a greater need for both IS and DA than would the mining company.

- Information is more strategic for some companies than for others:

  - For deregulated industries, information is more strategic.

  - It is particularly strategic in commercial manufacturing, where competition is worldwide.

  - It is less strategic in regulated industries and in natural monopolies. They cannot use information to gain an advantage over their competitors.

- Therefore, in general, DA is more valuable to enterprises that:

  - Are oriented toward masses of data.

  - Can use information to a competitive advantage.

- These criteria are summarized in Exhibit V-1.

EXHIBIT V-1

CRITERIA FOR ESTIMATING NEED FOR DATA ADMINISTRATION

STRONG NEED		WEAK NEED
Volatile Enterprise	versus	Stable Enterprise
Interactive	versus	Noninteractive
Large	versus	Small
Produce Masses of Data	versus	Produce Few Data
Data are Strategic	versus	Data Nonstrategic

VI HOW TO SET UP OR STRENGTHEN DATA ADMINISTRATION

VI HOW TO SET UP OR STRENGTHEN DATA ADMINISTRATION

- This chapter first presents an idealized description of how to establish a strong and effective DA function. Then more realistic methods are presented.

- There are two prerequisites to any method:

  - Getting the support of top management.

  - Getting the involvement of users.

A. THE PREREQUISITE STEPS

1. GETTING TOP MANAGEMENT SUPPORT

- Some respondents observed that "Without top management support, you don't really have DA. You just have a data dictionary."

- Proper establishment of DA requires a sequence of theoretical-appearing work. Without management's understanding of this work, support for it may wither.

- DA changes people's roles (as discussed in the next chapter). People are more likely to accept these changes if the DA effort enjoys firm public support from top management.

- How does DA get this support? The first step is to emphasize that management is rational.
  - Management wants to know what it is going to invest in.
  - Management wants to know what the benefits will be.
- The data administrator can give higher management an inventory of the possible benefits discussed in Chapter IV. These include:
  - Fewer problems with redundant data.
  - Better application programming.
    - Faster, higher quality system development.
    - Purchased software packages easier to install.
    - Better use of FGLs and DSSs.
  - Potentially improved management and executive information systems.
- Vendors may be biased regarding the value of their wares. However, a leading DA consultant, Dr. Robert Holland, said in 1982: "Companies today may expect a return of three dollars for every one dollar spent on the implementation of technology." Considering the return over a period of years, that estimate may be conservative.
- The following items are necessary to an estimate of costs:
  - First, approximate the labor hours required for:

- . Consultants, modelers, data designers, etc.

- . Interviews by modelers to get information about the essential jobs in the enterprise.

- . Record-keeping support.

- . Management support.

- Add the cost of any software tools that are purchased to aid in such functions as normalizing data models.

- Subtract the cost of any planning activities that will be replaced.

- The sum is the estimated net cost of establishing DA.

- Or, following the ideal method, use the following very rough rule of thumb:

  - Estimate the number of different kinds of essential, information-handling jobs in the enterprise.

  - Multiply that number by $2,000.

  - This product is the first rough estimate of cost.

- With management, frankly compare costs and benefits.

2. GETTING END-USER INVOLVEMENT

- Establishing or strengthening DA requires significant time and attention from end-users for two reasons:

- End users are some of the most immediate beneficiaries of data-oriented development. DA lets end users make better use of their PCs, DSSs, and FGLs.

- They must provide much of the requirements information that DA will need in designing different views of the shared data.

- They will gain a better understanding of the available data.

- They will come to appreciate the value of shared data - and ask for extensions of it.

- DA cannot simply demand the time of the end users. It can:

  - Request that they individually donate time to DA.

  - Request that their managers make end users' time available.

- Both requests should be made, out of respect for both end users and their managers. The request should be buttressed with clear descriptions of the benefits that they end users expect to receive from a broader DA function.

- DA and end users will be jointly looking at their requirements, identifying common requirements, and resolving conflicts. Therefore, end users with both interest and analytic ability should be the individuals whose help DA requests.

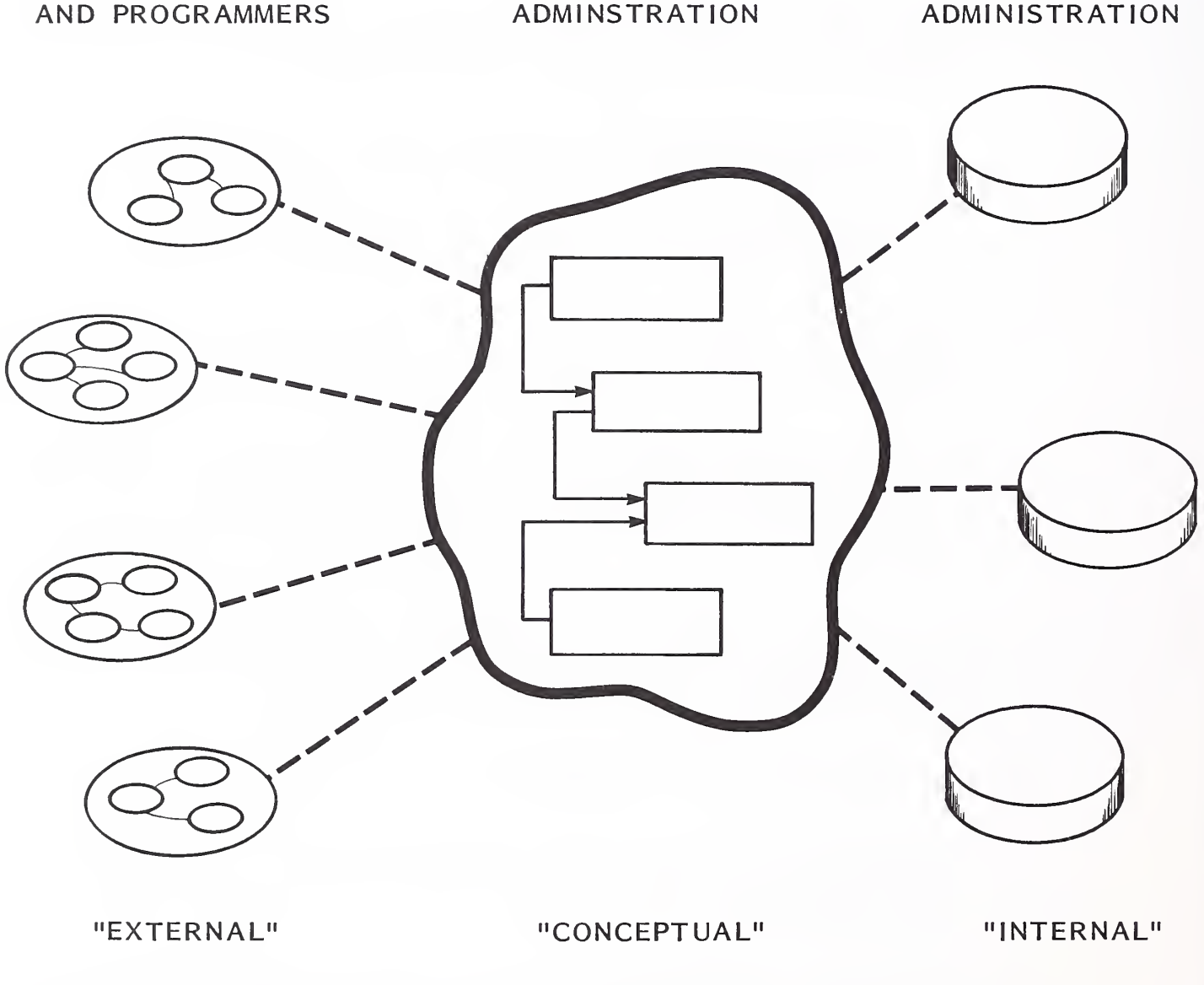
B. THE IDEAL METHOD: PREVIEW AND COMPROMISES

1. PREVIEW

- A basic concept behind the ideal method is the ANSI/X3/SPARC three-scheme architecture.
  - This architecture assumes that the data can be described in a single logical or conceptual data model (CDM). This CDM is:
    - Independent of any external application of the data.
    - Independent of any physical storage of the data.
  - Plural user views, showing how to tap available data, can be mapped onto the CDM. (These are frequently called external data models.)
  - Plural designs for integrated data bases, in the physical storage media, can be mapped to the CDM. (These are frequently called internal data models.)
  - An illustration of the three-scheme architecture, with notes on DA and DBA, is reproduced in Exhibit VI-1.
- The ideal method, in essence, consists of:
  - Identifying the functions of the enterprise to create a "business model" of the enterprise.
  - Organizing the information in the business model to produce a "normalized" CDM.

EXHIBIT VI-1

AN ANNOTATED ILLUSTRATION
OF THE THREE-SCHEME ARCHITECTURE



"EXTERNAL"

- Data Acquisition Software
- Data Application Software

"CONCEPTUAL"

- Integrated Logical Data Model
- Integrated Data Dictionary

"INTERNAL"

- Integrated Physical Data Bases

- Generating and documenting user views of the CDM.

- Generating and documenting physical views of the CDM.

- Using a DBMS-DD team to

  . Provide users with easy access to data.

  . Protect data integrity, security, etc.

- These activities, decribed in more detail in Section C below, are summarized in Exhibit VI-2.

2. EVIDENCE AND CONFLICTS

- Evidence regarding the value of the ideal method came from:

  - DA managers interviewed by INPUT.

    . Descriptions of what worked for them.

    . Descriptions of what they would like to do or like to have done.

  - The published literature.

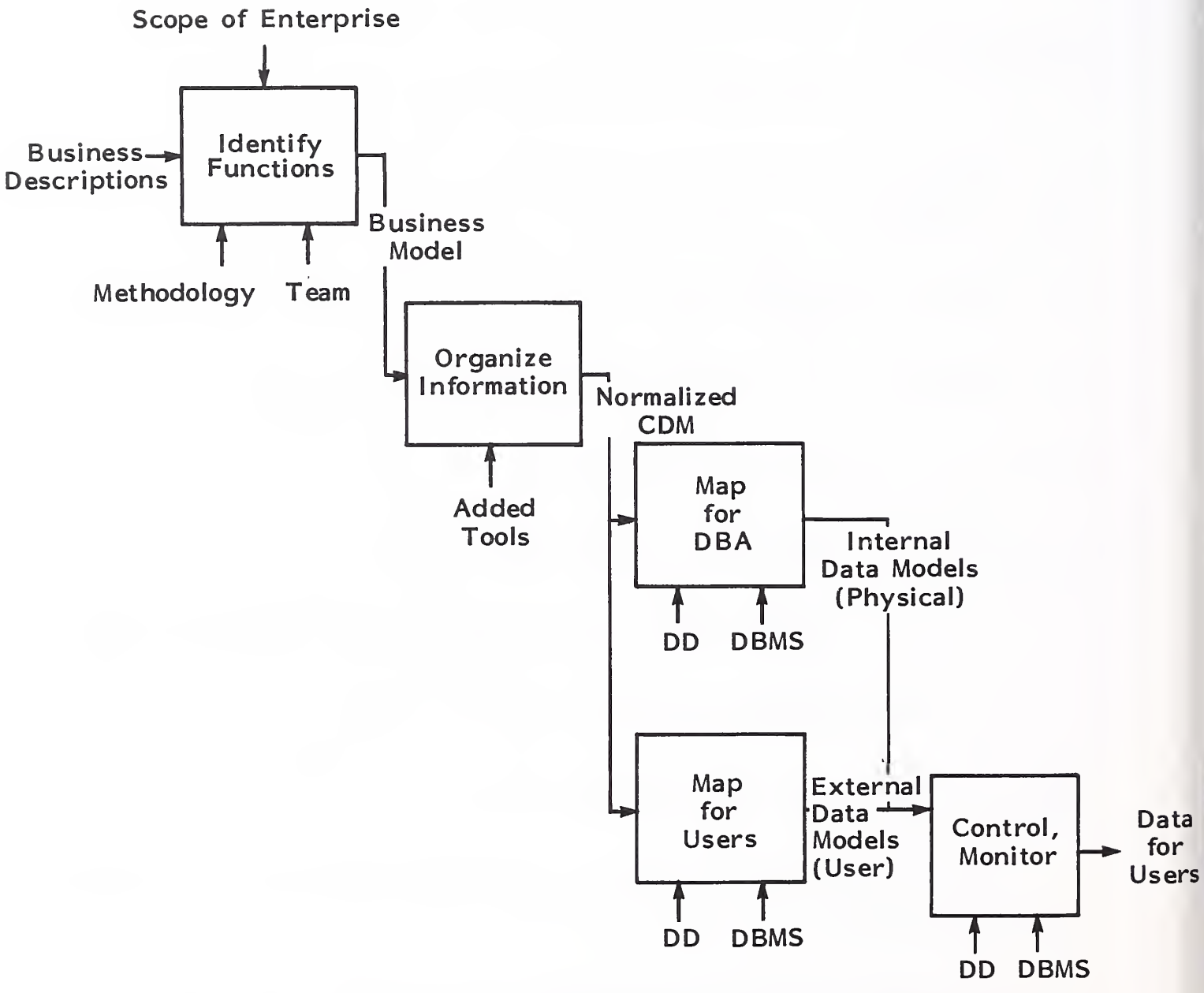
  - Academicians and researchers in DA.

- The ideal method conflicts with the recommendations of some of the vendors and consultants.

  - Some vendors bypass the business modeling step.

EXHIBIT VI-2

SUMMARY OF ACTIVITIES IN THE IDEAL METHOD



. They start with user views, then advocate that the user buy their tools to synthesize user views into a conceptual data base.

- There is a subtle danger in bypassing the business modeling step:

  . Existing user views probably do not represent all the uses that should be made of the corporate data resources. (For a manufacturing company cited in Chapter IV, identifying markets for replacement parts was an example of these new uses of the data resources.)

  . To the extent that profitable new uses are hidden, DA fails to live up to its strongest charter.

3. COMPROMISES

- Realistically, data administrators may not be able to get the authority to follow the ideal method with their entire enterprise.

  You can make at least three kinds of compromises with the ideal.

  - First, remember that there was no plan to do a data model of the whole corporation--just a model of a well-defined enterprise within the "whale." DA can try to carve out a smaller organ from the enterprise and, as a demonstration, apply the ideal method to it.

    . If there is a DSS center and a set of FGL users, start by trying to serve them. This would prove the value of the DA concept to end users.

    . It would also give you an indication of the magnitude of the implementation difficulties.

- Any incremental approach has the merit of avoiding massive conversions, which no one wants to undertake until they are absolutely necessary.

- Second, pressure for more automated enforcement of standards may be built up by concentrating on what one consultant called "the biggest failure of DP": Lack of discipline, rigor, and standards.

- Third, the data administrator can fall back on one of the less-than-ideal methods described in Section VI D and E.

C. THE IDEAL METHOD: SPECIFIC ACTIVITIES

- The activities described below will mutually overlap, both logically and temporally. They are separated below in a manner that gives maximum clarity.

1. GET A METHODOLOGY

- A business/information modeling methodology will be a way of developing a business model that will influence the data model and in turn the quality of the data base.

- The choice of methodology will be influenced by the way enterprise is viewed and can be perceived as:

  - A conventional structure, as in Exhibit VI-3.

  - A system contributing to the economic or social environment and feeding off that environment, as illustrated in Exhibit VI-4.

EXHIBIT VI-3

THE ENTERPRISE AS A STRUCTURE
(THE "X" AMPLIFIER COMPANY)

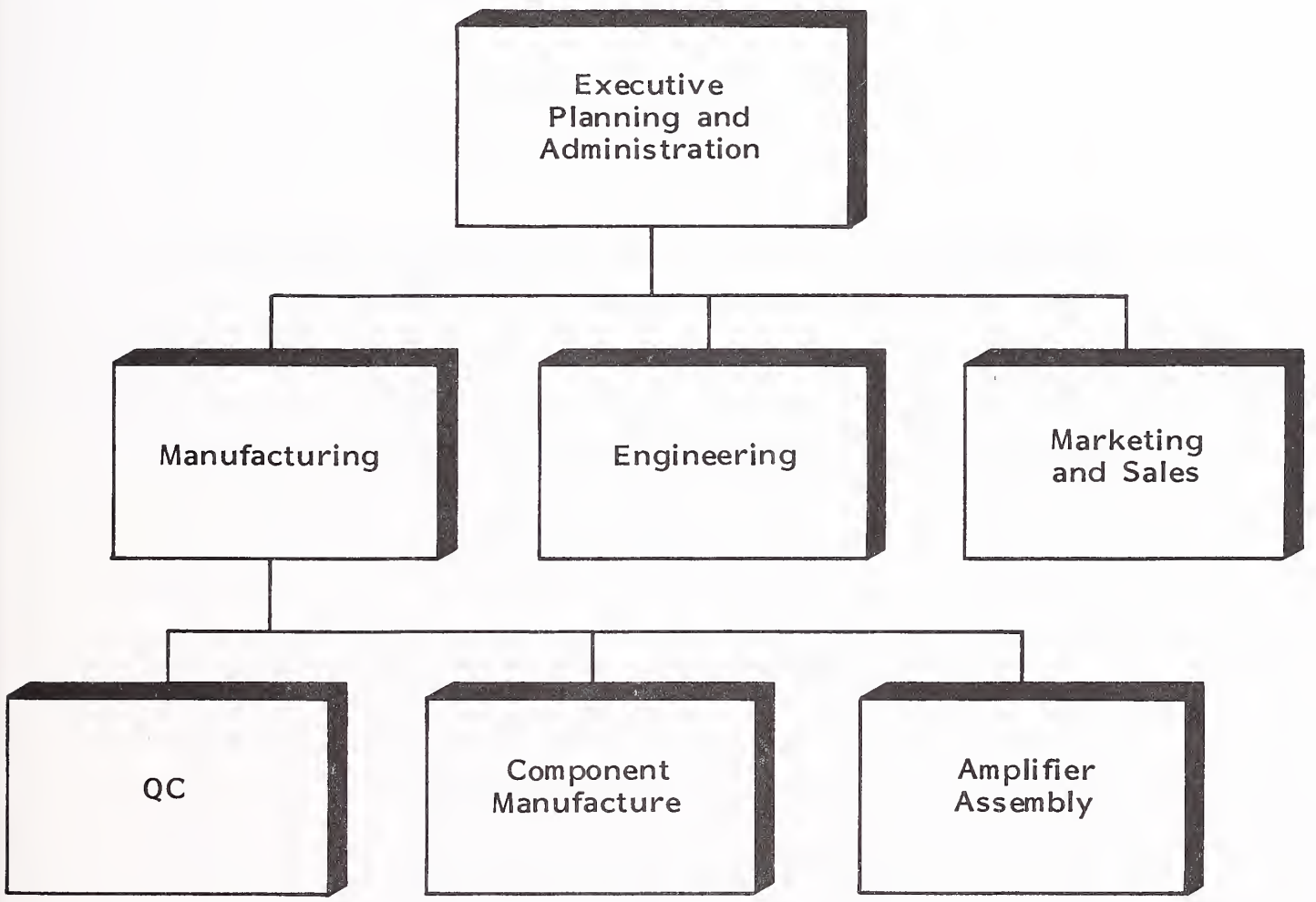
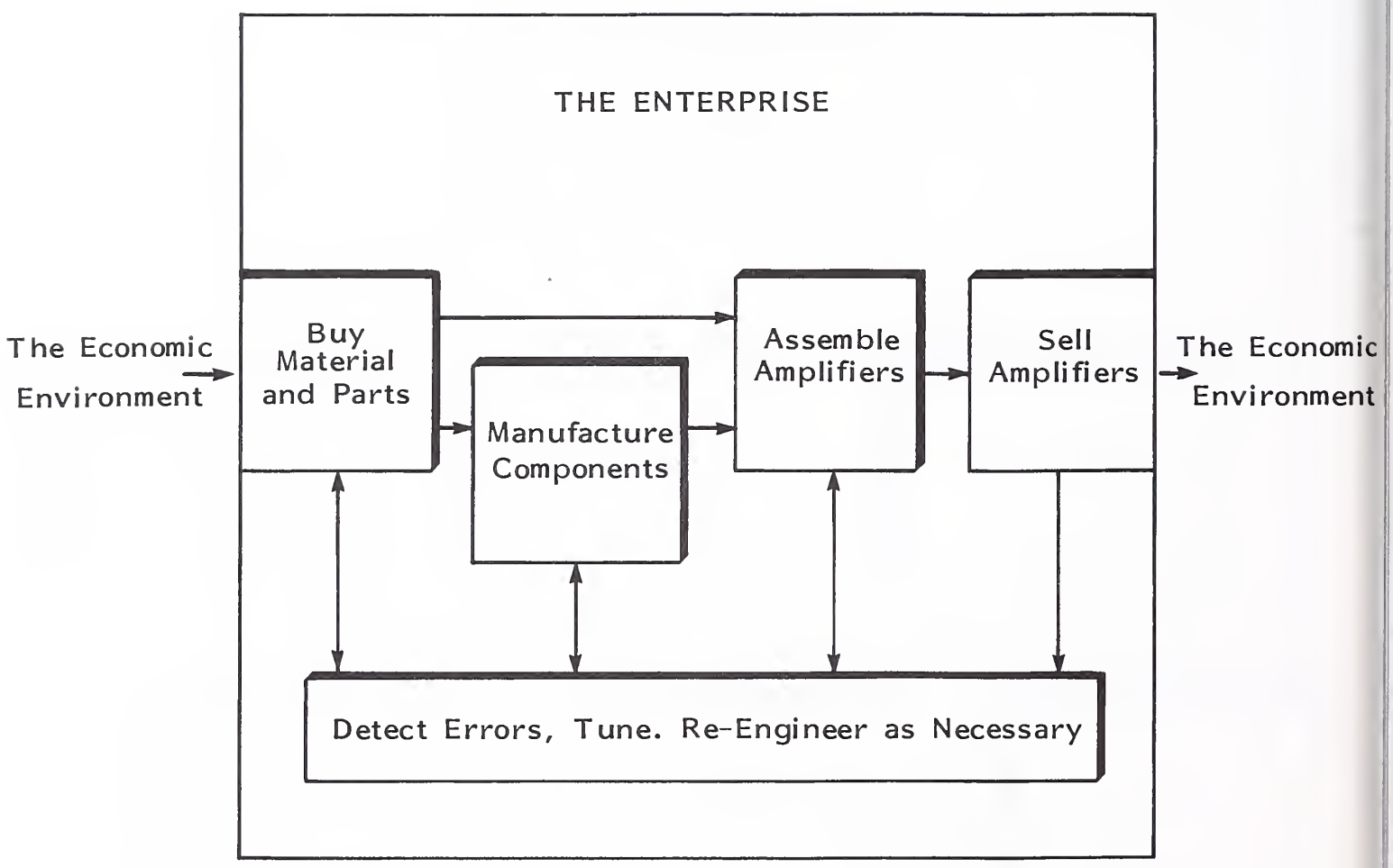


EXHIBIT VI-4

THE ENTERPRISE AS A SYSTEM
(THE "X" AMPLIFIER COMPANY)



- Neither view is wrong. The two views are just different perspectives of the same enterprise.

  - To the extent that your organization is viewed as a structure, it will be comfortable to use methodologies that:

    . Emphasize planning for the data base systems rather than prototyping them.

    . Are compatible with functional decomposition of requirements and emphasize requirements traceability (to know what software to change if business requirements change).

  - Most methodologies view organization as a structure, but are also compatible with the organization-as-a-system concept.

    . Prominent examples include IBM's Business Systems Planning (BSP) and James Martin's Data Planner (Database Design, Inc.).

    . There are many others.

  - To the extent that an organization is viewed as a system, managers may feel more at home using methodologies that:

    . Emphasize data base prototyping.

    . Are relatively simple and flexible.

  - For an analysis of planning methodologies, see INPUT's Integrating Systems and Corporate Planning, (March, 1984).

- Even if consultants guide the model building, IS and user personnel will have to work closely with them. Important criteria in selecting methodologies are the degrees to which people:

- Understand the methodologies.

- Feel they are realistic for the environment.

- The world has methodologies for everything from designing data bases to cooking eggplants. Even in the DA area, different methodologies start from different points of departure. DA methodologies may be based on:

  - Describing the enterprise and the data it needs (business modeling).

  - Reorganizing the data in homogeneous clusters (information modeling).

  - Designing software that will:

    . Store the data.

    . Use the data.

    . Identify which data and functions become obsolete if requirements change.

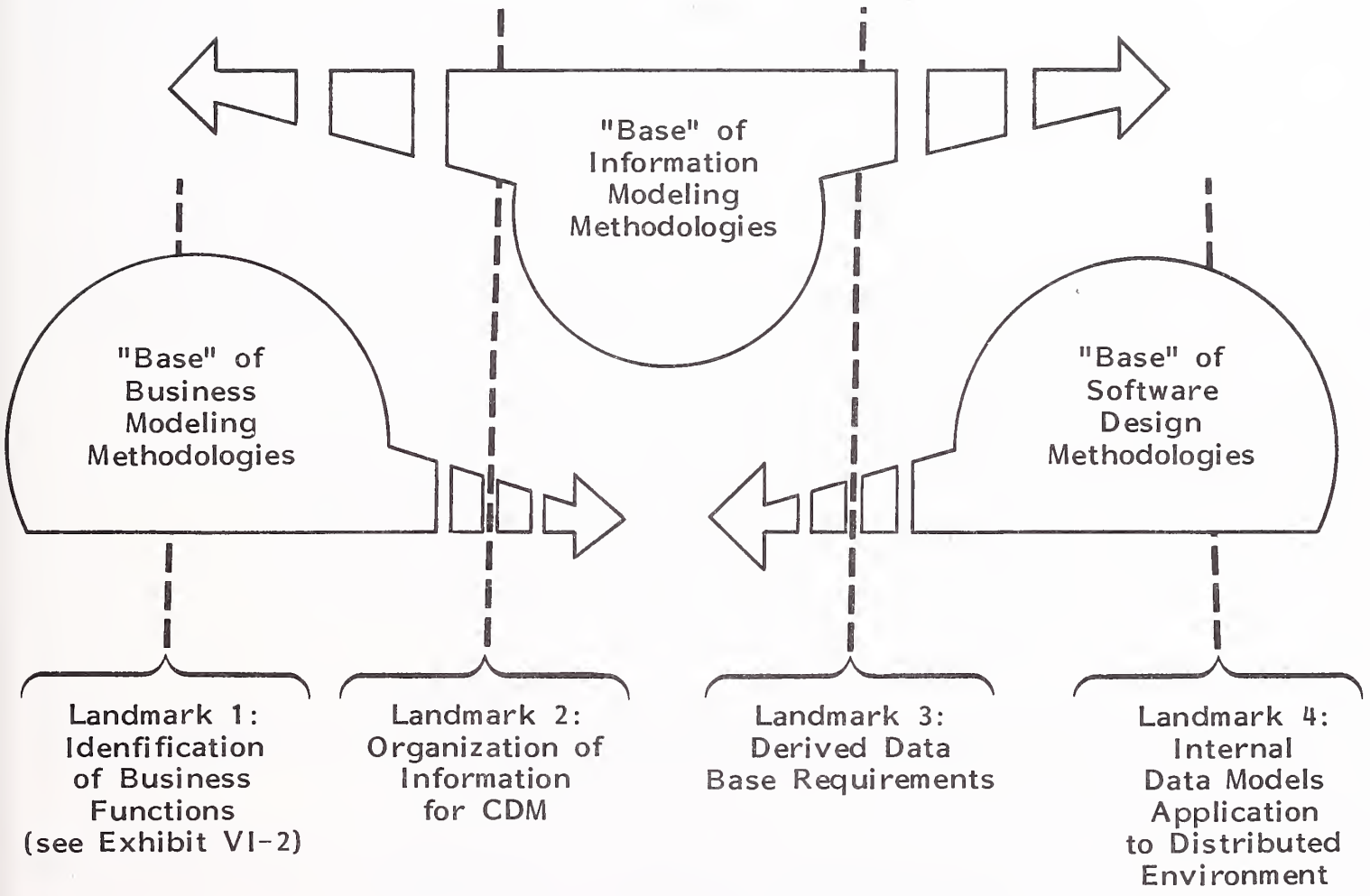
- From its point of departure (or "base"), each methodology covers a certain number (or "range") of related topics. Different "bases" and "ranges" of different methodologies are illustrated in Exhibit VI-5.

  - IBM's BSP, for example, takes business modeling as its point of departure.

    . BSP places strong emphasis on strategic objectives, business process models, finding out who uses what data, the identification of logical data areas, etc.

EXHIBIT VI-5

"BASES" AND "RANGES" OF METHODOLOGIES



  - . Less emphasis on data architecture, derivation of logical system structures, etc.

  - . No emphasis on physical data base design, applicability to a distributed environment, etc.

- Other methodologies have comparable limitations. This is inevitable.

  - . The subject matter changes from business processes and interactions to data models and data use.

  - . The details of the methodology have to change along with the subject.

- Note that the purpose of any business/information modeling methodology is to produce a model that is:

  - Adequate in detail.

  - Accurate.

  - Clear.

- If the model is clear, its results can be used by other methodologies as attention moves away from business and toward data.

  - The disadvantage of a change in methodologies is that it would require extra effort in translation of results.

  - The advantages are:

    - . The new methodology should be better for the purpose at hand.

. IS would not be trapped into depending upon a single methodology consultant or vendor.

- In summary, INPUT recommends that IS should:

  - Choose a methodology with which IS feels comfortable.

  - Demand clear results.

  - Be willing to change to a new methodology or consultant as the work changes from describing the business to organizing data structures.

2. RECRUIT YOUR TEAM

- In practice, this step will start shortly before selecting a methodology.

  - Talk with different methodologists: your own (if your enterprise has them) and consultants. Decide what combination of methodologies and methodologists seems most appropriate to your enterprise. Not only do different consultant prefer their own methodologies, but their application background areas differ. They may have worked in:

    . Manufacturing.

    . Finance.

    . Utilities.

    . Software development.

    . Academic research.

    . Etc.

- Other things being equal, pick a methodologist who knows your industry.

- Your team should consist of:

  - The person who is going to become the data administrator. They should be a generalist and not necessarily from IS.

  - An expert in modeling of:

    . First, the business and its functions.

    . Later, the data and information.

  - People who know the subject being modeled, but who don't necessarily know how to model.

    . First, people who know your business.

    . Then, IS experts.

  - A librarian or secretary.

  - Reviewers, to check the models for accuracy and logic.

  - The IS manager or sponsor of DA, who is frequently talking with top management and explaining what this foolishness is all about.

3. CREATE A BUSINESS MODEL

- The true concept of DA is that information is an enterprise-wide resource that should be exploited and therefore managed and controlled. Accordingly,

DA requires a basic understanding of the very fabric of the business and the environment in which it operates.

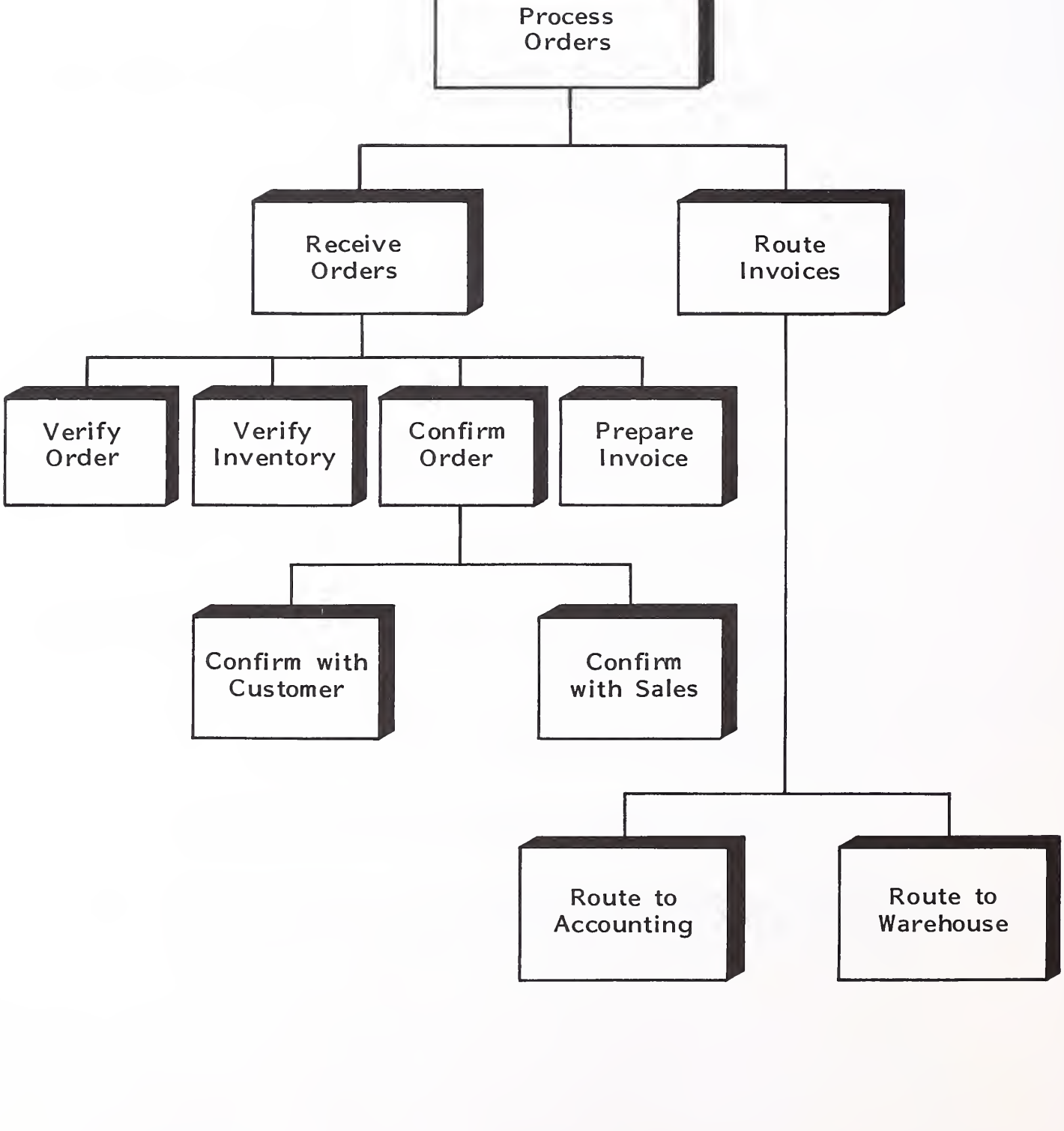
- The business model will represent that basic understanding.
  - Business functions are identified in clear, concise terms.
  - "Entities," or the essential things in the enterprise, are then indentified.
- Exhibit VI-6 shows a simplified business function model for a small parts distribution enterprise.

4. DEVELOP A CONCEPTUAL DATA MODEL (CDM)

- This step:
  - Isolates the entities that are most essentially involved in the functions in the business model.
    - A customer may be an entity.
    - An order may be an entity.
  - Indicates which entities have some kind of direct relationship with other entities. For example:
    - A customer places an order.
    - Inventory satisfies an order.
- A CDM can be graphical, tabular, or both.

EXHIBIT VI-6

SIMPLIFIED BUSINESS FUNCTION MODEL
FOR PARTS DISTRIBUTOR



- Exhibit VI-7 shows the essential beginning of a CDM for the small parts distribution enterprise.

  - The model is "conceptual" in that no consideration has yet been given to physical storage of the data.

  - The labeling in Exhibit VI-7 conforms to the government-sponsored $IDEF_1$ language. Because it is widely used in Industrial Modernization Incentive Programs and related contracts, $IDEF_1$ is the most standard language for information modeling.

  - The $IDEF_1$ language and associated methodologies are in the public domain (available only to U.S. citizens). Information about them may be obtained from the Integrated Computer-Aided Manufacturing Program Office, AFWAL/MLTC, Building 653, Wright Patterson AFB, OH 45433.

  - In Exhibit VI-7, the labeling in the bottom left corners of the entity boxes is an $IDEF_1$ convention. It leaves room for notation of attributes above the entities.

- An attribute is a "value" that goes with an entity; e.g., an order has a dollar value, or inventory holds a certain number of entities in stock.

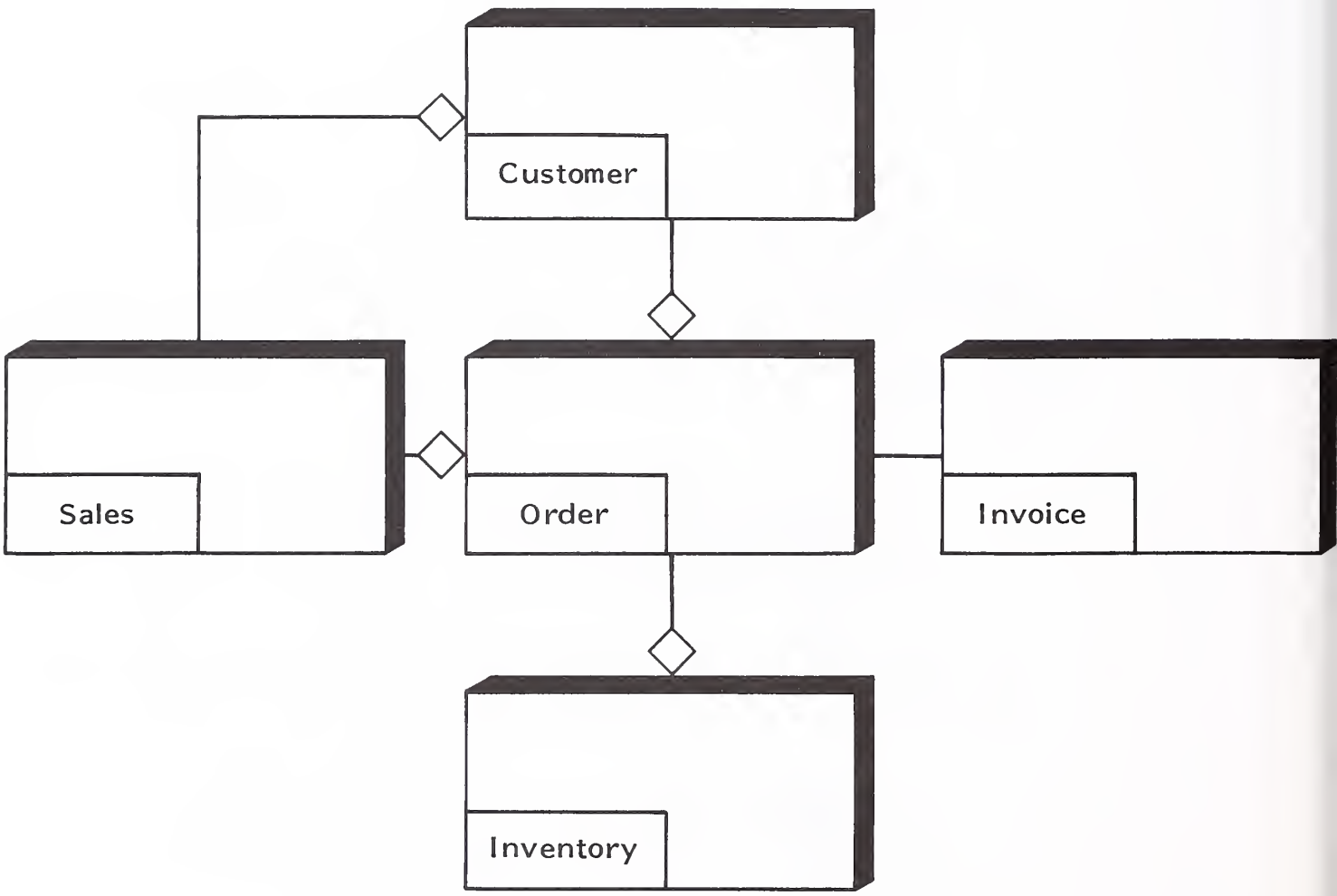
  - An attribute's value may be a classification; e.g., the entity EMPLOYEE may have an attribute called GENDER, and the "value" of GENDER may be MALE or FEMALE.

  - An entity typically has several attributes; e.g., pay rate, hire date, gender, etc.

- The little diamonds by some of the boxes in Exhibit VI-7 indicate plural relationships (e.g., one customer may place more than one order.

EXHIBIT VI-7

ESSENTIAL ELEMENTS OF CDM FOR PARTS DISTRIBUTOR



Models with plural-to-plural relationships are normally redone (or "normalized," as described below) to minimize such relationships (because they are ambiguous).

- Exhibit VI-8 shows a table relating the entities and functions for the parts distributor. The table maps the functions in Exhibit VI-6 onto the entities in Exhibit VI-7. Such a table is called a "data matrix."

  . A data matrix table is not theoretically necessary.

  . However, it will greatly facilitate loading the DD.

- Warning! The CDM must be capable of growth and change.

  - The functions of the enterprise may change.

  - The things or "mechanisms" performing the function should change.

    . The DA concept should make it easier to computerize new functions.

    . Robots may replace people in manufacturing.

  - Data bases developed by the CDM approach are more adaptable to changes in mechanisms than are those developed by older approaches. The CDM does not intermix the mechanism and the information. Therefore, changes in mechanisms will not effect the information.

- "Normalizing" the model increases its capacity to grow and change without becoming ineffective or too awkward.

EXHIBIT VI-8

DATA MATRIX

FUNCTIONS	ENTITIES				
	Sales	Customer	Order	Inventory	Invoice
Verify Order		X	X		
Verify Inventory			X	X	
Confirm Order-Customer		X	X		
Confirm Order-Sales	X		X[1]		
Prepare Invoice		X	X		X
Route Invoice-Accounts					X
Route Invoice-Sales					X

5. NORMALIZE THE CDM

- Normalizing the CDM rebuilds it in terms or entities that are usually more generic. For example, HOURLY EMPLOYEE YEAR-TO-DATE VACATION PAY might become EMPLOYEE CLASS TIME PERIOD PAY TYPE.

- The terms or entities also tend to become more cohesive or homogeneous. For example, all entities associated with EMPLOYEE would be more directly linked together in a normalized CDM.

- The structure of the normalized CDM is also more logical. For example, in a normalized CDM, GENDER could not exist without an employee first existing.

- The normalized CDM is also more complete (and therefore looks more complicated).

- Normalization carries strong advantages:

  - Data bases derived from normalized models are more expandable. More information can be added to them without changing their basic structure.

  - Logical access routes are more apparent and more limited in number.

  - Maintenance is easier. Changes can be made to a class of entities rather than to one entity at a time.

  - Searches are more complete.

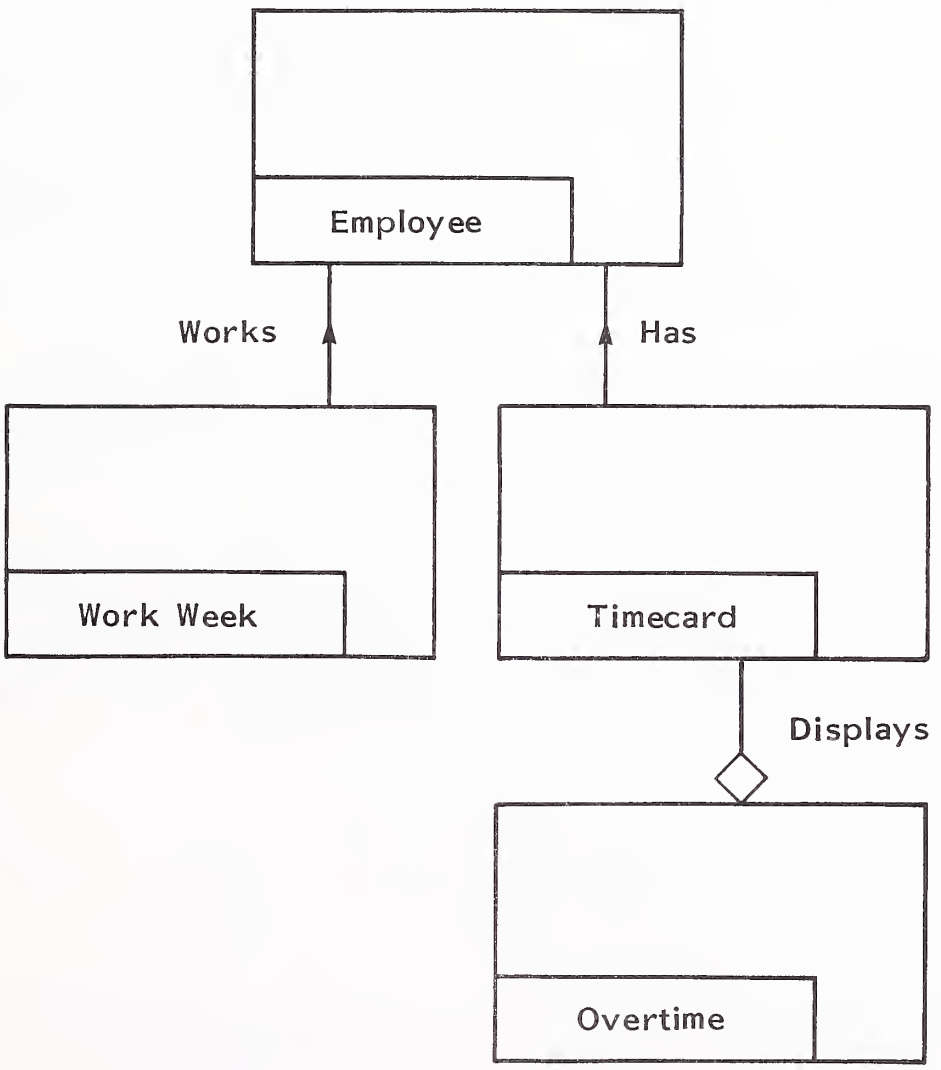
    . As mentioned in the example below, every individual entity will have every attribute of every member of that class of entities.

- . This completeness of attributes ensures that searches on the basis of attributes are complete.

- There are different degrees of normalization.
  - Each degree "buys" more of the above advantages, but costs more in design time.
  - These degrees are measured in "normal forms."
    - . Because future expansion needs are hard to predict, it is difficult to say what degree of normalization a CDM will need.
    - . Different consultants advocate everything from the second to the fifth normal form, and this has been satisfactory to its users. INPUT recommends that CDM developers plan on going to the third or fourth normal form.
- An example from the IDEF$_1$ methodology will illustrate what happens to a model as it goes through repeated iterations. (These accomplish the equivalent of normalizing the model.)
  - Exhibit VI-9 shows an early-stage IDEF$_1$ model of a payroll system.
    - . No attributes (like GENDER) have been entered in the entity boxes.
    - . However, the modeler has decided that OVERTIME is an entity in its own right, and not an attribute of TIMECARD or EMPLOYEE.
    - . The reason is that, for every class of entities (e.g., EMPLOYEES), there should be a class of attributes (e.g., OVER-

EXHIBIT VI-9

EARLY STAGE IDEF$_1$ MODEL



TIME), and each individual entity (i.e., "entity value" or "type") should have every attribute; but not all employees are paid OVERTIME (i.e., salaried employees are not).

- Exhibit VI-10 shows an expansion of the model to include the various accounts to which time is charged. This model is "nonspecific" because it includes a "many-to-many" relationship: Many timecards may show the same account, and many accounts may be shown on the same timecard.

- Exhibit VI-11 shows the removal of the nonspecific relationship.

  . Each timecard may display many charge numbers (or maybe just one).

  . Each account appears as many charge numbers.

  . The new entity, CHARGE, connects timecards with accounts.

- Now the modeler decides to add entities whose existence depends on the CHARGE. That is, OVERTIME and REGULAR TIME could not exist if CHARGE did not exist. The resulting model appears in Exhibit VI-12.

  . "Dependencies" are marked by the heavy arrows embedded in some of the lines in Exhibits VI-9 through VI-12.

  . Identification of dependencies helps generate generic entities such as EMPLOYEE CLASS rather than HOURLY EMPLOYEE, and PAY TIME rather than VACATION PAY.

- As the configuration of entities begins to stabilize, classes of attributes are added to the boxes. Additions are shown in Exhibit VI-13.

EXHIBIT VI-10

$IDEF_1$ MODEL WITH NONSPECIFIC RELATIONSHIP

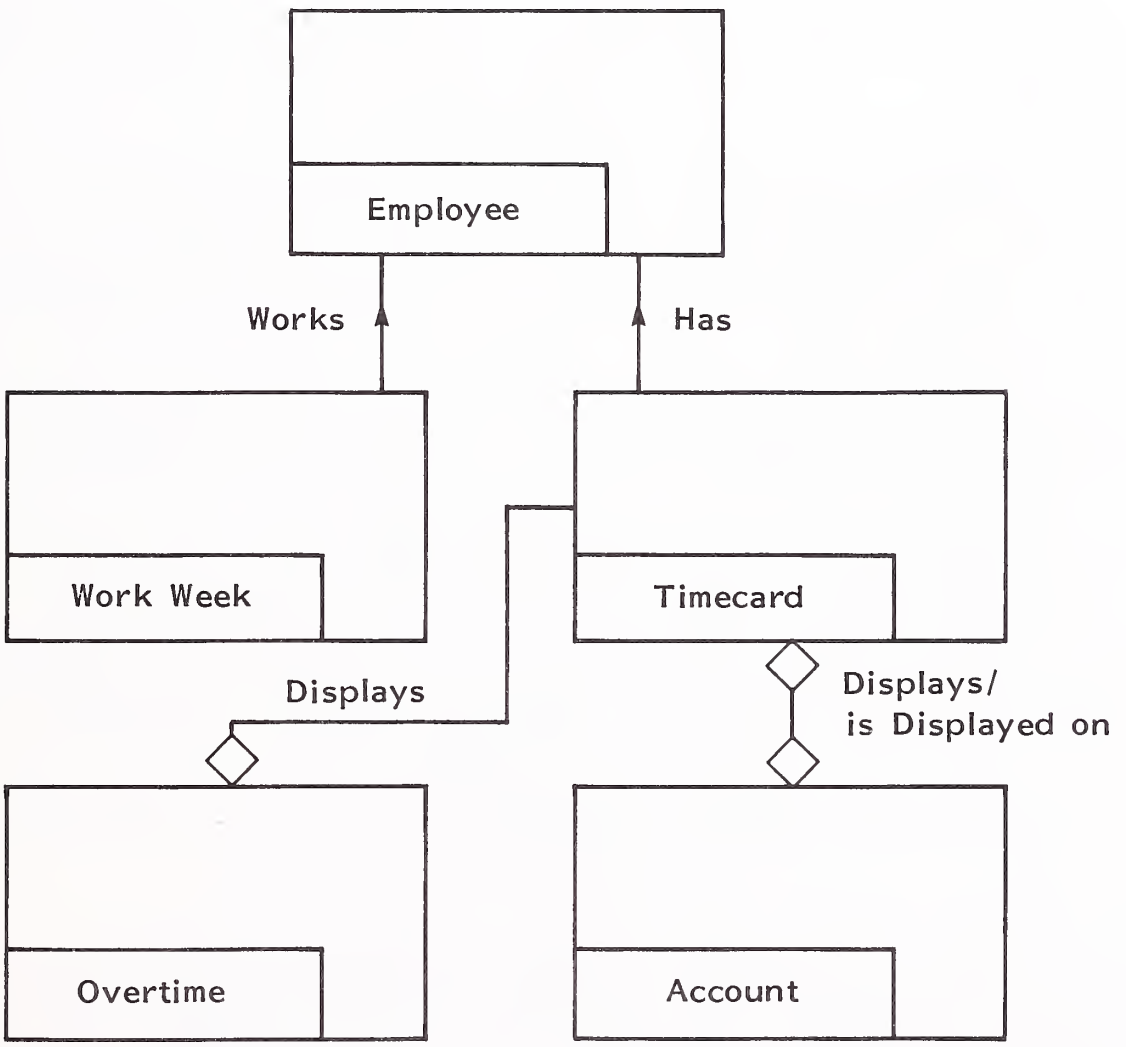


EXHIBIT VI-11

ADDITION OF ENTITY TO
REMOVE NONSPECIFIC RELATIONSHIP

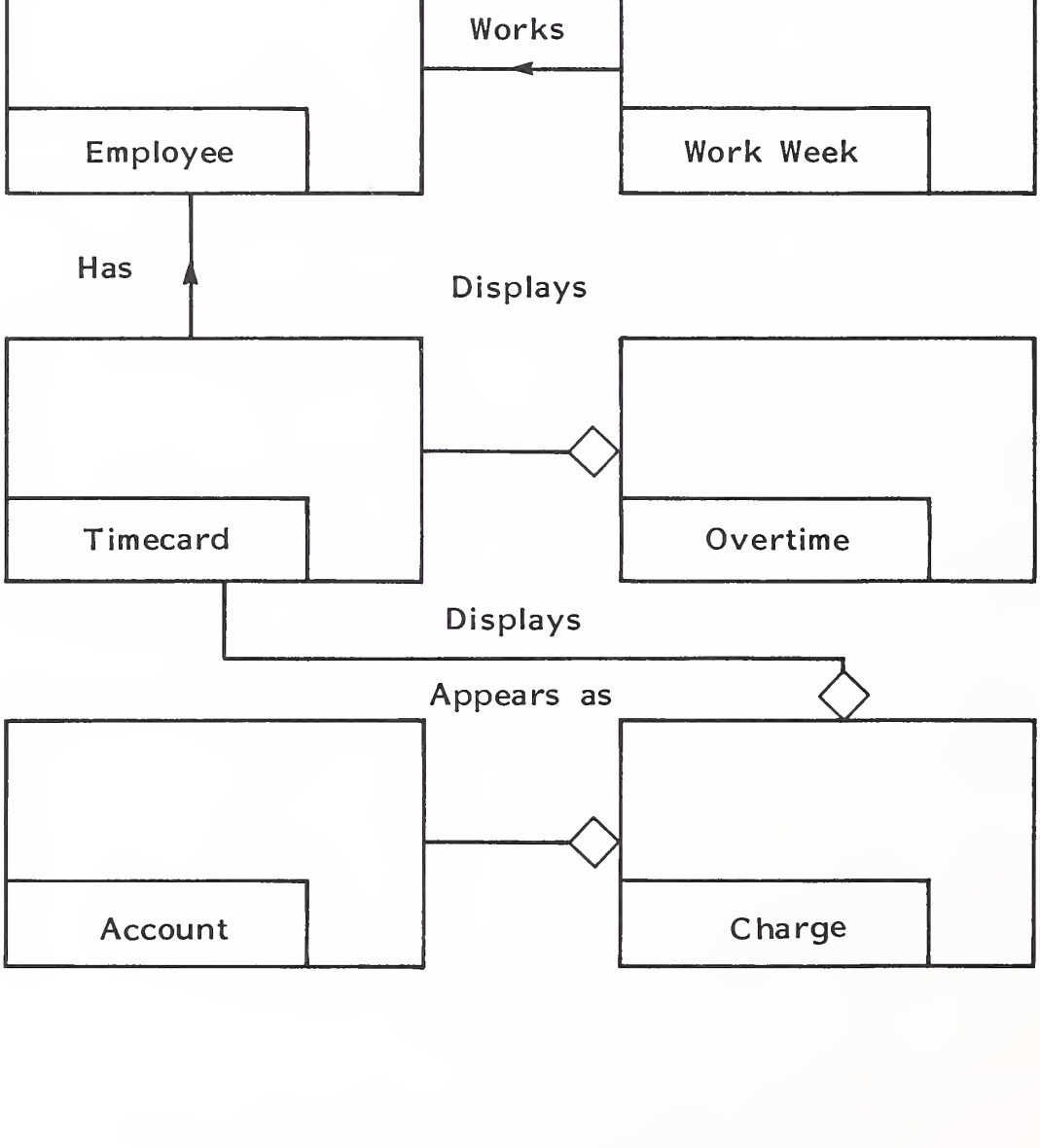


EXHIBIT VI-12

REFINEMENT OF MODEL - ADDING DEPENDENT ENTITIES

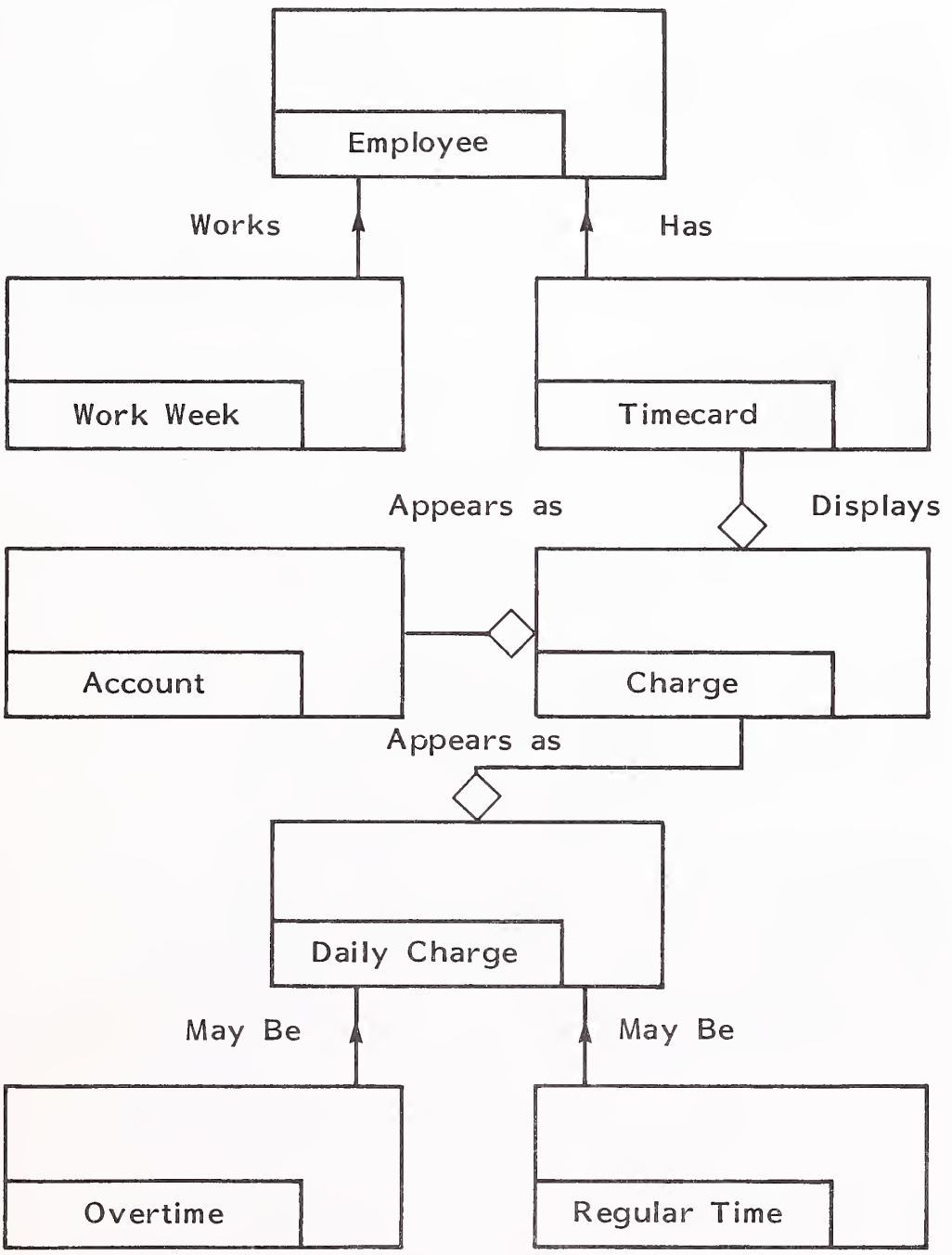
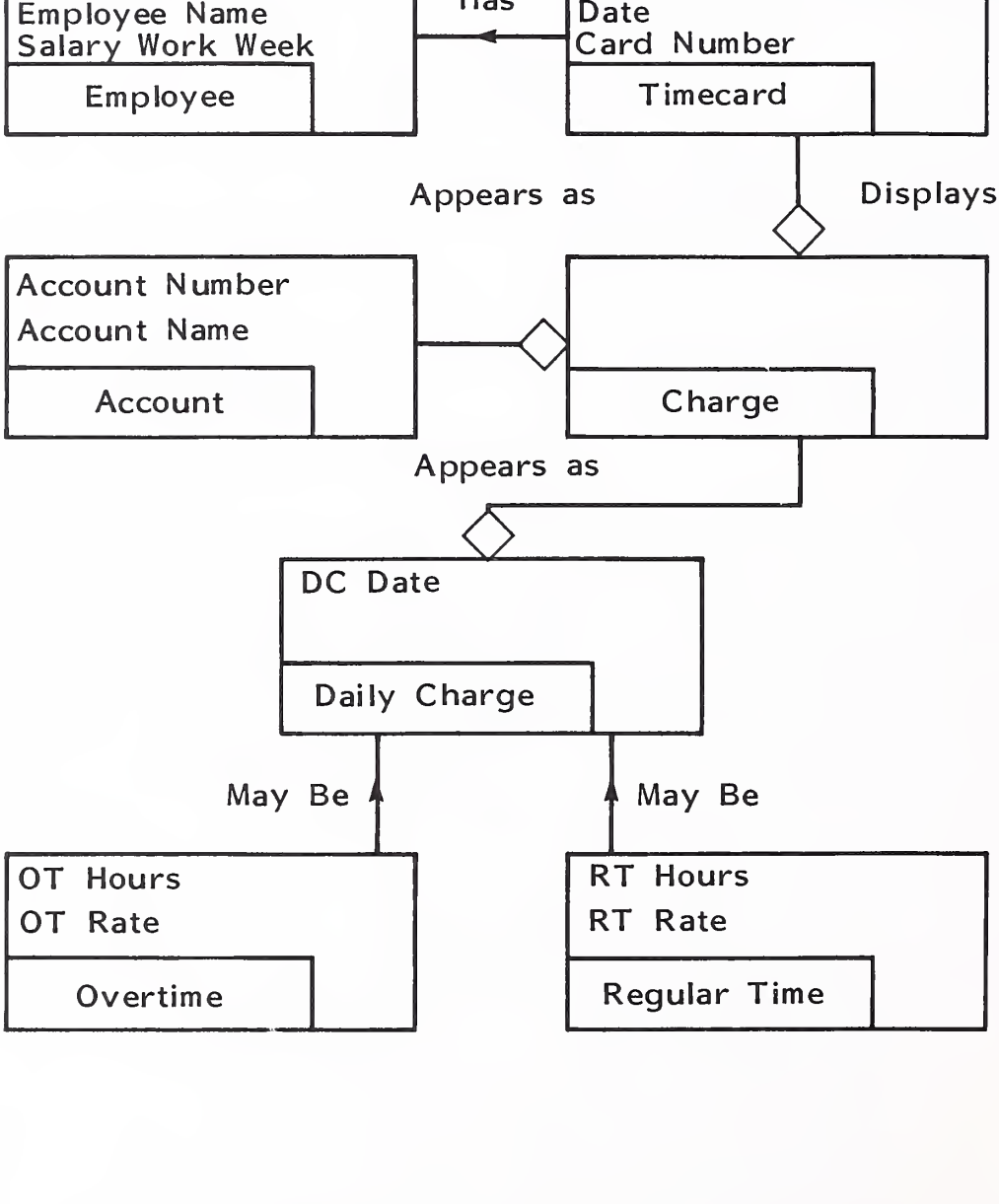


EXHIBIT VI-13

$IDEF_1$ MODEL WITH ATTRIBUTE CLASSES ADDED



- Also, dependent entities are examined to see if they are really attributes. In Exhibit VI-13, WORK WEEK turns out to be an attribute of EMPLOYEE because:

  . WORK WEEK must exist for each EMPLOYEE.

  . No employee can simultaneously have two WORK WEEKS.

  . OVERTIME and REGULAR TIME remain entities because they carry other attributes (i.e., HOURS and RATE) that could not be represented if OVERTIME and REGULAR TIME were attributes.

- The data in all of the attribute classes can convey useful information (e.g., pay rates).

- Some attribute classes convey other information. They are "key classes."

  . They are sufficient to identify entities.

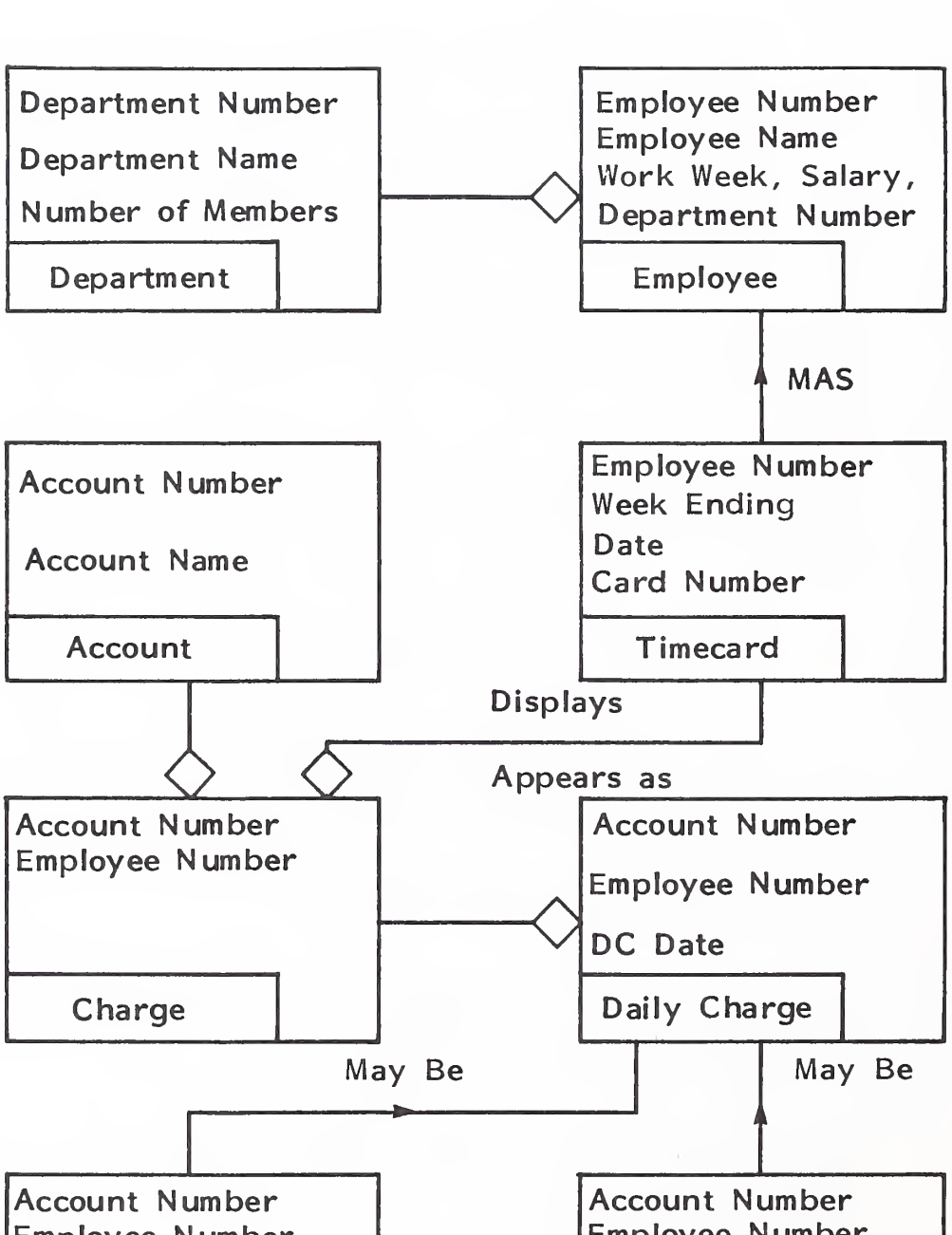
  . For example, NAME might not be sufficient to identify an employee. (There might be two Mary Jones.) But EMPLOYEE NUMBER would uniquely identify any member of the class of all employees of the enterprise.

- If an attribute is a key class, it is underlined in the $IDEF_1$ language. This is illustrated in Exhibit VI-14.

- Key classes "migrate" down. Dependent entities must have the key class of the "parent" entity.

EXHIBIT VI-14

$IDEF_1$ MODEL WITH KEY CLASSES IDENTIFIED



  - For example, in Exhibit VI-14, TIMECARD is dependent on EMPLOYEE for its existence, and TIMECARD must, like its "parent," be identified by EMPLOYEE NUMBER.

  - Key classes also migrate as ordinary attributes. For example, DEPT # becomes a nonkey attribute of each employee in that department.

- Finally, synonyms and local definitions may be added to the model by "tags." For example, Exhibit VI-15 shows that:

  - SOURCE CODE is sometimes used as a synonym for DEPT #.

  - In some parts of the enterprise, ACCOUNT # is called CHARGE #.

  - Such tags will later become important parts of the data dictionary.

- Exhibits VI-9 through VI-15 indicate that building a large normalized data model would be a lot of work. This is true.

- An early decision should be made: Whether to buy software tools to aid in building and normalizing the models. INPUT's recommendation is:

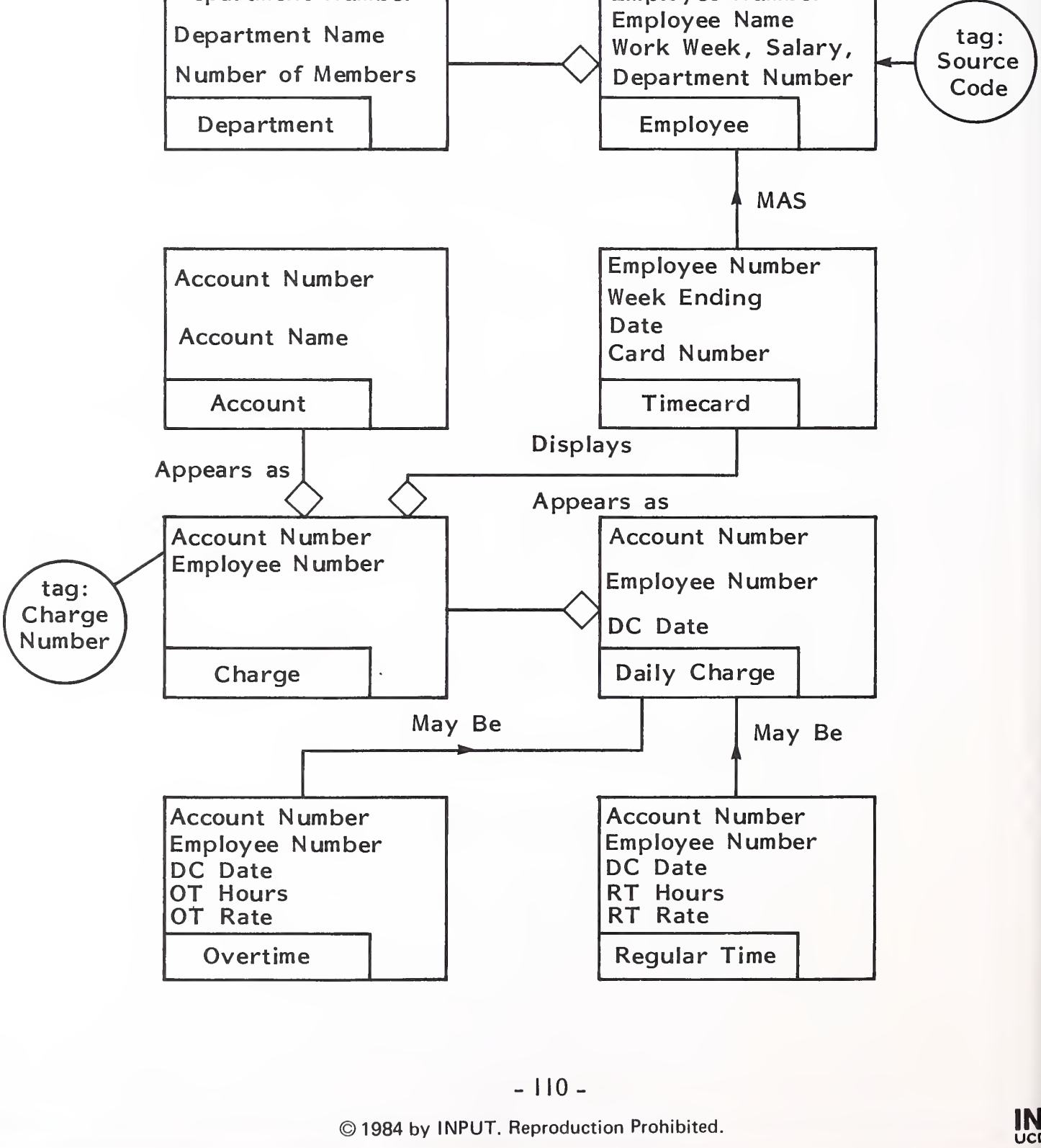
  - Get an estimate (from your own experts and/or consultants) of the time needed to build and normalize the model of your enterprise. Include:

    - The time of the experts and consultants.

    - The time of other personnel.

EXHIBIT VI-15

FINAL IDEF$_1$ MODEL WITH ADDED TAGS



- If the value of the time is more than $30,000, start shopping for software tools. Vendors include Database Design, Inc. and Holland Systems.

6. GET AND START LOADING A DATA DICTIONARY (DD)

- The DD will serve as the repository of all information about the enterprise's data resources.

  - It will show the way to the data.

  - It will provide the "matrix" through which data can be interpreted as information.

- The team will load the DD with (at a minimum):

  - Entity/relationship/attribute definitions.

    . Where definitions differ (as in the "tags" on an $IDEF_1$ model), the differences should be noted.

    . The English-language parts of the definitions should make business sense to the users.

    . The DD should contain the normalized attributes in the entity records.

  - Entity/relationship/attribute uses.

    . These identify the various user areas that have an interest in the entities.

- They also specify the systems, programs, files, and records that use the data elements.

- Application programmers and other users should be able to see the definitions in the DD.
  - Access should be easy for the user to accomplish.
  - To protect the data, user access should be on a ready-only basis.
- Note that the ideal approach is to get a DD before acquiring a DBMS.
  - In practice, the opposite has generally happened.
  - The criteria that should be used in selecting a DD are:
    - Capability of handling the definitions (discussed above) and the data models or views (discussed below).
    - Capability of providing both user access and data security.
    - Capability of interacting with the DBMS (discussed below).
    - Capability of controlling use of the DBMS, rather than being just a documentation system.
- Choice of the DD is extremely important in minimizing conversion problems.
  - DDs usually allow some flexibility in naming standards.
  - However, all have some degree of structure/inflexibility in this respect.

- If a lot of money has been invested in a DBMS, examine the vendors' DDs closely to see if they will mesh with the DBMS with little conversion work.

- In the ideal approach, compare the DD's flexibility with the range of entries from your models.

7. DEVELOP AND DOCUMENT INTERNAL DATA MODELS

- Given the CDM, data base designers can map out efficient physical structures for the data base. These are the internal models, or views, of the data. They should be:

  - Shareable.

  - Nonredundant.

  - Independent of the physical structure of the data base.

- Data Base Administration (DBA) should be responsible for the generation and maintenance of the internal models.

  - The design expertise should reside with the Data Base Administrator.

  - The Data Base Administrator will normally bear the daily, operational responsibilities.

- The Data Base Administrator's organization should be responsible for documentation of the internal models.

  - DA should be exercising a regulatory and control function.

  - DA should have control of the DD.

8. DEVELOP AND DOCUMENT EXTERNAL DATA MODELS

- These will provide individual program or user views of the data for:
  - Data maintenance.
  - Information retrieval, via
    - Appropriate access paths.
    - Special query and similar languages.

9. GET A DBMS

- The DBMS is the mechanism for converting the internal and external data models into a usable form.
  - The selected DBMS should be capable of handling the size and complexity of those models.
  - It should also provide access paths to the data element definitions in the DD.
- The DBMS must be compatible with an "active" DD; i.e., the DBMS must obtain its definitions from the DD. This arrangement:
  - Reduces labor in maintaining definitions.
  - Automatically enforces naming conventions and other standards.
- The DA concept requires basic units of shareable data. These, as provided by the DD and DBMS, are the accessible entity records. These records:

- Represent homogeneous subdivisions of data and define specific subjects of interest to different people in the enterprise.

- Will be in a normalized form, to improve shareability.

- Homogeneous, normalized data records provide a pure, solid foundation for developing higher levels of information for moving up the pyramid illustrated in Exhibit III-9.

D. THE EVOLUTIONARY APPROACH

- In the absence of at least a test version of the ideal method, a more "natural" or "evolutionary" approach is recommended. The Data Administrator should try to guide the evolution through four stages:

  - Initiation.

  - Expansion.

  - Formalization.

  - Maturity.

- The administrators should decide where they are on this evolutionary scale, and then try to strengthen their position from that point.

- Typically, a DBMS is already in place in the enterprise.

1. INITIATION

- If Data Administrators are in this phase, they should:

- Seek maximum commitment from top management, as discussed in the beginning of this chapter.

- Establish a charter for DA, using as strong a definition (from Chapter III) as possible.

- Select and install a DD. Plan to make the DA a control mechanism (as described in the ideal method) as soon as possible.

2. EXPANSION

- In the expansion phase, the administrator should seek additional line responsibilities, including:

  - An active role in file and data base design.

  - Direct control over the DD, including

    . Security.

    . Defining data, etc.

  - Documentation of more systems in the DD.

3. FORMALIZATION

- In this stage, the Data Administrator should centralize the following under DA:

  - Data definition authority.

  - Data base design expertise.

- In a large company it is logical for:
  - DA to become a department in face and in name.
  - DA to acquire more formal management controls.
  - More application systems and data bases to be brought under DD service and control.
- By this time it should be possible to demonstrate clearly:
  - The lack of a long-range systems architecture.
  - The need for a global system based on a unifying concept such as the three-scheme architecture.

4. MATURITY

- In the maturity stage:
  - Many of the master files and data bases used in the enterprise are documented in the DD.
  - So are most data elements, some of which will have been established as standard definitions.
- Now Data Administration can:
  - Continue the above documentation and standardization of definitions.
  - Cooperate in the development of hierarchies of information that will satisfy top management.

E. THE DATA DICTIONARY APPROACH

- This is the most common and least desirable approach.
- It typically develops through these steps:
  - An organization starts using data base concepts, and probably a DBMS.
  - The organization realizes it is not fully exploiting the potential of data base systems. It sees a lack of:
  - Standards regarding data.
  - Control and planning of data in general.
  - In an attempt to induce standardization, a DD is brought in.
  - The DD is used merely as a cataloging tool. But the cataloging does not produce desirable results.
  - The DD is not actively coupled to the DBMS.
  - There is no CDM to unify the data as a manageable resource; i.e., there is no organizing mechanism and the result is a mess.
- In this situation, the DA manager has at least two choices:
  - Move into the initiation phase of the evolutionary approach, and start to use the DD as a control mechanism.
  - Define an enterprise and advocate the ideal approach for it.

- In any case, the Data Administration must realize that problems in cataloging and other areas are evidence of a bad approach. It is as if one were creating an ordinary dictionary for which an alphabet had not yet been invented.

F. LOCATION OF DATA ADMINISTRATION IN THE COMPANY

- Where should DA fit within the organizational structure of the company? There is a general consensus on these points:

  - Perhaps DA should report to the same person to whom the head of DP reports, but DA should not be subservient to System Development or part of IS. The goals of IS and DA are opposed:

    . IS wants to develop systems as quickly as possible.

    . DA wants to guarantee data quality and is not interested in "putting out fires."

  - DA needs to satisfy the end users, and therefore DA should be in a position in the organization that is visible to end users.

  - DA should be in a high enough position to have some regulatory and planning authority.

  - Although DBA has been around longer than DA, DA should not report to DBA.

  - DA should be in a position to interact with the company's planning and control organization.

- Three possible positions for DA are analyzed below.

I. REPORTING TO TOP MANAGEMENT

- DA is a staff function within the corporate office.

- This is the preferred position, one that offers the maximum leverage for achieving corporate data/information resource management objectives.

  - The corporate perspective gives a better view of the information requirements of the entire organization.

  - Because of the high level of this position, DA is likely to receive more cooperation and acceptance.

- Some respondents saw a need for a Chief Information Officer (CIO) who reports to the chief executive officer.

  - One person, predicting CIOs in the future, quoted James Martin: "information is the competitive edge of man's future, just as the wheel was the competitive edge of his history."

  - A position paper of a major utility company with a progressive DA function declares:

    . "By the end of the decade, information will play such an important role in the company that it should begin to appear as an asset on corporate financial reports.

    . "As the role of information increases, consideration should be given to the creation of a Chief Corporate Information Officer. . ."

- If these predictions are valid, there will be a trend toward Data Administrators who reports to top management.

2. REPORTING TO THE CONTROLLER

- This arrangement breaks DA away from the IS department. However, DA is not a department or staff function in its own right.

  - This arrangement, like the above, has the advantage of providing a corporate perspective on the information resources.

  - DA gains some "muscle" by its association with the financial authority of the controller's office.

3. REPORTING TO I.S.

- This arrangement implies two opinions of DA.

  - First, it is a way of reorienting IS toward greater user satisfaction.

  - Second, it is a responsibility center for ensuring quality in system development efforts.

- This arrangement creates a narrow perspective on data resource management.

  - It creates the danger that users will be alienated by the misdirected control imposed through the data dictionary.

  - It offers little opportunity to move toward directly satisfying some of the information needs of higher management.

- This is the least desirable position.

G. SUMMARY

- This chapter has told how to attain certain goals through certain activities.
  - The goals are the benefits resulting from a CDM and integrated, expandable data bases:
    - . End-user computing that is both easier for the user and better controlled.
    - . More productive professional programming.
    - . The basis for an executive information system.
  - The activities are both human and technical. They include:
    - . Getting top management support.
    - . Generating a business model that will in turn lead to a CDM for the enterprise.
    - . Normalizing the CDM to lay the groundwork for integrated, expandable data bases.
    - . Using a DD and DBMS, partly to minimize conversion problems.
    - . Actually developing the physical data bases.
    - . Developing views of these data bases that will best serve the users.

- If not all the above activities are possible, some compromise approaches to good DA can be taken.

- Finally, good DA deserves a high position in the hierarchy of the enterprise. At a minimum, it should report to some kind of planning and control organization.

VII MAINTENANCE AND HUMAN ISSUES

VII MAINTENANCE AND HUMAN ISSUES

A. CHANGING THE CULTURE

- Respondents frequently talked about DA's need to "change the culture" in which it worked. But what is a culture?

  - A culture is a set of roles.

    . In a primitive culture, roles include animal skinners and cleaners, witch doctors, and chieftains.

    . In an IS culture, roles include data entry clerks, business and scientific programmers, and IS managers.

  - A culture is a set of values.

    . In a primitive culture, magical ability is valued.

    . In an IS culture, programming ability is valued.

  - People in a culture are threatened by anything that will change their roles and values. Technology makes such changes.

    . Modern medicine threatens witch doctors and downgrades their role.

- Data base technology threatens to change roles and values in the IS culture.

- The professional histories of three representative careers in an IS culture are outlined in Exhibit VII-1. If DA infuses modern data base technology into their culture, their roles and values will change.

  - Business programmers:

    - May feel that their expertise in structured COBOL is being rendered obsolete because others perform well with FGLs.

    - May feel insulted that people are now talking about Customer Data Bases rather than Accounts Receivable.

  - Scientifc programmers:

    - May look down on those who do not understand the exciting package concepts in Ada and instead fuss over dull data bases.

    - May prefer the elegance of C to the "hairiness" of large data bases.

  - The system development managers:

    - May suspect that their system development expertise is not appreciated.

    - May wonder whom they can manage if the people they currently know how to manage are replaced.

EXHIBIT VII-1

A TRIO OF TYPES IN THE I.S. CULTURE

BUSINESS PROGRAMMER	SCIENTIFIC PROGRAMMER	I.S. MANAGER
Started Programming in COBOL	Started Programming in FORTRAN	Started Programming in Assembly Language
Wrote an Accounting Package	Wrote a Fast Fourier Transform	Wrote a Matrix Inversion Package
Now Working on an Insurance Claims System	Now Working on an Embedded Program for a Blood Tester	Now Working as System Development Manager
Now Writing in COBOL	Now Writing in C	Now Writing in English
Takes Pride in Knowledge of Accounts Receivable	Takes Pride in Logical Ability	Takes Pride in Ability to Manage Programmers
Loves Structured COBOL	Has a Love-Hate Relationship with Ada	Loves Predictable Working Relationships
Feels Threatened by Data-Oriented Development	Feels Contemptuous of Data-Oriented Development	Feels Insecure with Data-Oriented Development

- The concerns of these people are to an extent realistic. DA will not allay their concerns by denying this fact.

- A better approach is to point out history: New technology always creates new opportunities as well as displacements.

  - Business programmers can apply their knowledge of Accounts Receivable to the design of user views of the CDM, and these views will make the programmer and others more productive in the future.

  - Scientific programmers may find the mathematics of normalization as interesting as those of transform theory--and, in some businesses, more valuable.

  - IS managers may be able to apply their system development and managerial talents to helping a larger group of predictably naive end users.

- A time of change is always a time of testing. If the DA concept is seriously implemented, then there will be real changes and tests. In general, the change will affect:

  - End users.

  - Other technicians as well as the programmers.

  - Other managers as well as the head of system development.

- If DA has the charter or willingness to do so, it should deal with these changes at two levels:

  - Politically, at the managerial level.

- Through training at the lower levels.

- The word "political" is used in its good sense: Working out agreements for the good of the community.

  - Conscientious establishment of the DA concept will probably require changes in the organizational structure. A good political tactic is for the DA manager to confer with his or her director and jointly work out a tentative new organization chart. Then, in deciding to what extent to implement that chart, let the DA leader take the technical and consulting role while the director takes the executive and political role.

  - Training has two advantages.

    . For those who are trained, it promises to solve their problems--to give them new or expanded skills.

    . It is not emotional. In most progressive companies, occasional training sessions are expected, and a training department is set up to facilitate those sessions.

  - Here is a tip for the training department: Training in data base technology is based on specific technical facts. It is not like leadership training; it is not based on psychology.

  - Instructional System Development (ISD) is an effective and well-established methodology that is available to teach technical skills. It is well known and is documented, among other places, in Air Force Manuals 50-2 and 50-8.

    . If IS is approached by training consultants who don't know what ISD means, it shouldn't hire them.

- Finally, going back to Chapter VI, there is tremendous value in firm, open support from upper management.

B. MAINTENANCE

- Stability is perhaps the greatest single appeal of the data model. Data types (e.g., defect rates, orders, hours of work) tend to be more stable than the mechanisms or processes that create these data.

- If a data model is properly built (i.e., adequately normalized), it should change very little until the company gets a new type of business, or until the structure of the enterprise changes significantly.

- Still, the data model is a snapshot (albeit an expensive snapshot) of the enterprise, and the enterprise does change. All data model managers surveyed by INPUT reported at least some maintenance activity. With that activity comes problems.

  - The biggest problem is human. People tend to neglect maintenance.

    . Documentation is part of the infrastructure, and it is unglamorous and seems never urgent. So, under pressure, the infrastructure deteriorates: People neglect maintenance in favor of "putting out fires." (The analogy is apt: Does the crew of a fire truck stop and paint hydrants on the way to a blazing house?)

    . Perhaps because of the lack of glamour, few people seem interested in doing maintenance work.

  - A related problem is technical. A data model is complicated.

- A doctoral respondent, who is interested in artificial intelligence, commented that both data models and "knowledge-based systems" (in artificial intelligence) are "fragile." That is, both require a skillful maintainer. A few mistakes (e.g., with key attribute classes) can do great damage to the structure that enables the data to convey useful information to people.

- The most creative people, however, prefer design work to maintenance work.

- If maintenance of an application program is neglected, people are afraid to use the program and don't trust its documentation. The same fate can befall data models.

- DA managers have a choice.

  - They can budget perhaps half a person and a software tool for maintenance.

  - Or they can neglect maintenance and risk losing the base of the system.

C. TOOLS AND CONSULTANTS

- A "new pressure" regarding DA was explained in Chapter III: Data will be shared, because it requires work to generate and input data. DA can either fight the data-sharing trend, or facilitate and co-opt it. (Here co-opt means "join a movement and then take it over.")

- DA can co-opt the data-sharing movement if it can offer superior wares that users will use. The need is to satisfy the users.

  - People like to play with tools.

    . FGLs are useful tools. DA should encourage their use.

    . Because some FGLs are linked to specific DBMS packages, the FGL should be a consideration in the selection of a DBMS.

  - People don't like to follow standards.

    . So, where possible, the DD should automate standards so that the user never sees the standards.

    . Clear standards must be written for the user. Sources of clarity are cited under Human Factors Engineering, below.

- Consultants were mentioned previously in connection with purely technical roles: providing expertise (e.g., on modeling or normalization) that the company does not have. However, consultants can also be valuable because they are unbiased with respect to the company.

  - Consultants have their own biases, which are very strong.

    . They want to promote their own ideas.

    . They want to sell their own services.

  - But they typically have no biases regarding political issues in the company.

- Therefore consultants are often very effective in explaining a new project.

- People will listen to them.

- The older consultants can recount histories of similar projects at other companies.

- INPUT recommends the limited use of consultants as a way of clarifying the interaction between the technical and the human aspects of DA.

D. HUMAN FACTORS ENGINEERING

- Human Factors Engineering (HFE) is an established discipline. Much of "I" involves the physical design of workspaces, control devices, etc., and is not relevant to DA. But parts of HFE are relevant, and they should be used. These include:

  - Designing the "display" to be visible to the eye and understandable to the mind. This part of HFE sometimes has relevance to DA. Some of its principles and findings concern:

    . Error rates in scanning digits.

    . Parts of video displays that are most likely to be seen.

    . Changes that are most and least likely to be noticed.

    . The understandability of labels and messages.

    . When "information overload" occurs.

- Making the procedure compatible with human nature. This concerns the effects of sequencing human actions in different ways. For example:

  . If a person hears a number and says it to someone else, he/she is less likely to make a mistake and also less likely to remember the number.

  . If a person hears a number and writes it to someone else, he/she is more likely to make a mistake but also more likely to remember the number.

- If DA gets involved in designing any kind of complex format that must be understood by a variety of users, HFE could probably help in the design process.

VIII RECOMMENDATIONS FOR A DATA ADMINISTRATION STRATEGY

VIII RECOMMENDATIONS FOR A DATA ADMINISTRATION STRATEGY

A. PREMISES

- DA, buttressed by data-oriented development, constitutes a wave of the future.
  - End-user computing is proliferating. This causes a dilemma.
    - Much greater responsiveness to user needs is possible.
    - Without an adequate DA function, much end-user computing will (a) be based on inaccurate data, (b) degrade the accuracy and security of corporate data, or (c) both use and create bad data.
  - DA offers the immediate promise of significantly:
    - Increasing the speed with which IS produces new application systems.
    - Raising the quality of these systems.
  - DA offers the future promise of contributing to better executive information systems.
- DA will significantly change the structure of IS organizations.

- The scope of application development will narrow as "consumer computing" produces small systems. IS will concentrate more on large application systems and operating systems.

- As DA is brought in to IS organizations, DA will grow. DA will be perceived as encroaching on IS turf.

B. RECOMMENDED STRATEGY

- Three strategic questions must be answered.
  - In a given organization, in what areas should DA be first implemented?
  - What are the human considerations in implementation?
  - What is the best technical approach?
- The first implementation should be performed in an organization or sub-organization that is:
  - Large enough to allow demonstration of benefits, but small enough to minimize risks.
  - Integrated, in the sense that it would benefit from an integrated data base system.
- Human considerations extend to the entire organizational pyramid.
  - Some technicians will need new training to retain their value to the company.

- Satisfaction of end-users will be a major goal of DA. If end-users are not satisfied with DA's products, they will either not use or misuse corporate data.

- DA will change the organizational structure. A frank approach to these changes will minimize covert politicking.

- Without real top management support, DA is likely to fail. It is recommended that DA report to a high-level planning and control function, or that it be on the corporate staff.

- The technical heart of implementation is the three-scheme architecture through which a conceptual data model can be generated.

  - Prior development of a business model will minimize the risk of overlooking vital areas of business information.

  - It is essential that the conceptual data model be normalized if the later physical data bases are to provide the benefits of integration described under Premises (Section A).

  - Enforcement of data standards is necessary if the data bases are to be maintained and retain their value. An active data dictionary is recommended to automatically enforce standards.

C. CONCLUSIONS

- This report has covered the technical and human problems in establishing a good DA function, as well as the values of such a function.

- The technical problems are the easiest. Effective technical approaches are known, and one can buy the talent to implement these approaches.

- The human problems can be minimized by:

  . Explaining clearly to top management what is being done and why.

  . Collaboration in the planning of the organizational changes that any new technology will entail.

  . Using established principles of training and human factors engineering.

- There are positive values to good DA, and negative values to inadequate DA.

  . Without adequate DA, and organization faces increasing problems with data that are redundant, inaccurate, or unavailable.

  . With good DA, the evidence points to better application programming, better "consumer computing," and more truly informative management information systems.

APPENDIX A: USER QUESTIONNAIRE

USER QUESTIONNAIRE

I. First, let's try to define some terms.

1. How do you define Data Administration (DA)?
2. What are the most important other concepts (e.g., Information Resources Management, Data Base Administration) from which DA should be distinguished?
3. How do you define data modeling or information modeling?
4. Which of the above are you most interested in?
5. Why?

II. Now . . . some questions about data modeling.

1. What tools support it?
2. How well?
3. What techniques . . . ?
4. How well?
5. What tools and techniques are missing?
6. Have you been involved in data modeling, or are you familiar with examples?
7. In each case, what was the scope of the effort and the costs?
8. In general, where do you think data modeling should and should not be employed?
9. Can you name any good consultants in data modeling?
10. After someone builds a big data model, is it perishable? What information do you have on the costs of maintaining it?

III. Returning to DA in general, what do you think is the most promising:

1. Tools?
2. Methodologies?
3. Consultants?
4. Can you specify any need for integration of tools, (to make them work better as a set, or of tools and methods?

IV. Now . . . some practical questions about DA.

1. How is your DA function organized? How big is it?
2. What services does it offer to your DP community?
3. What restrictions does it have?
4. To what extent is DA in your corporation centralized?
5. What kinds of environments or enterprises need highly centralized DA functions?
6. Why?

V. "Data-oriented development" is widely advocated these days.

1. Do you favor it over "procedure-oriented development"?
2. Why?
3. What tools and techniques support it?

VI. Now we'll discuss some trends or factors affecting DA.

1. What trends concern or interest you most?
2. How are you dealing with these concerns?
3. In what industries and environments are your concerns most and least shared?
4. (If not already mentioned . . .) Is the influx of personal computers a concern to DA?
5. How, and to what extent, should PCs come under the influence of DA?
6. Do you have a concern about the use of external commercial data bases?
7. What should DA's role be regarding use?

VII. Now . . . some "bottom line" questions.

1. To what extent is good DA a technical versus a management issue?
2. Can you give us any example of bad DA caused by inadequate support of the DA function?
3. What were the results?
4. Can you give us examples of bad DA caused by poor strategy or poor practices?
5. What were the results?
6. What are some examples of good DA?
7. What were the results?
8. Looking at the basic business or service of your corporation, how would you say that DA ultimately affects its productivity?

VIII. If someone gave you a few million dollars and ordered you to retire, what advice regarding DA would you give to your replacement?

IX. What else should we have asked you but didn't?

APPENDIX B: VENDOR QUESTIONNAIRE

VENDOR QUESTIONNAIRE

1. There's a group of terms related to Data Administration (DA):

   IRM, Information Management, Data Base Adminstration, Information Engineering, etc. What term do you favor for the state-of-the-art data adminstration function, and why?

2. Do you supply tools or methodology consultants for data modeling or information modeling? If so, please tell us about them.

3. What, as a rule of thumb, is the cost of a data modeling effort?

4. For what kinds of clients is it most and least valuable, and why?

5. Can you give any case histories of payoffs to your clients?

6. Who are your major competitors?

7. What is the estimated cost of maintaining a data model?

8. What are the most valuable available tools for DA in general?

9. Methodologies?

10. Do you see any need for integrated sets of tools?

11. Consider three or four of your clients. Can you tell us about the organization and responsibility of DA in each of them? If the organization, etc., is not ideal, tell us how it ought to be changed.

12. Do you supply tools and consultants to support data-oriented development? If so, tell us about them briefly. If so, what data can you give us on the benefits your clients obtain from them?

13. Again thinking of three or four different kinds (sizes and industries) of clients, please tell us what trends or factors affect their DA function most.

14. If you did not discuss it above, tell us the desired and actual effects on DA of the following:

    (a) The influx of personal computers
    (b) Access to external, commercial data base services

15. Being as quantitative as possible, what case histories can you give of good DA functions that help organizations, and of inadequate ones that make organizations less competitive?

16. If you were talking to the CEO of a prospective client, what would you tell him/her about the way DA ultimately affects profitability?

17. If a friend of yours had just been made responsible for DA in a large company, what advice would you give him or her?

18. What else should we have asked you?